

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:38:26 ; Search time 73.69 Seconds  
(without alignments)  
22.610 Million cell updates/sec

Title: US-09-613-092a-7  
Perfect score: 79  
Sequence: 1 LVRRFVHRPHEVQ 15

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	15	20	AAV30353
2	74	93.7	15	20	AAV30354
3	52	65.8	156	22	AAV75292
4	52	65.8	156	22	AAV75293
5	52	65.8	243	22	ABBS8399
6	52	65.8	273	22	ABG18986
7	52	65.8	280	22	ABG10923
8	52	65.8	282	22	ABG10922
9	52	65.8	351	20	AAV59757
10	48	60.8	142	22	ABG18985
11	45	57.0	580	20	AAV34665

12	45	57.0	798	22	ABB62356	Drosophila melanog
13	44	55.7	164	22	AAH83043	Human immune/haema
14	42	53.2	175	22	AAV30377	C glutamicum prote
15	42	53.2	203	22	AAV78972	C. glutamicum SRT
16	42	53.2	316	22	AAV58773	Propionibacterium
17	41	51.9	105	22	AAV64657	Propionibacterium
18	41	51.9	466	19	AAW84355	Talaromyces thermo
19	41	51.9	466	20	AAV39899	T. thermophilus ph
20	40	50.6	1233	20	AAV55954	Mouse Str20-relate
21	40	50.6	1317	22	AAH65167	Drosophila melanog
22	39	49.4	63	21	AAV02947	Human secreted pro
23	39	49.4	76	22	AAV58639	Propionibacterium
24	39	49.4	194	22	AAV25661	G protein-coupled
25	39	49.4	409	22	ABB65172	Drosophila melanog
26	39	49.4	986	20	AAV37617	Protein which is s
27	39	49.4	1792	22	ABB71108	Drosophila melanog
28	38.5	48.7	830	22	ABG28908	Novel human diagno
29	38.5	48.7	973	22	ABG24942	Novel human diagno
30	38	48.1	67	20	AAV29052	T. gondii immunoge
31	38	48.1	67	22	AAV25523	Drosophila melanog
32	38	48.1	141	22	AAH69191	Propionibacterium
33	38	48.1	174	22	AAV52994	Propionibacterium
34	38	48.1	177	22	AAV42368	Human polypeptide
35	38	48.1	217	22	ABB65332	Drosophila melanog
36	38	48.1	275	22	AAV62724	Propionibacterium
37	38	48.1	312	22	AAV71996	Human olfactory re
38	38	48.1	312	22	AAV72164	Human olfactory re
39	38	48.1	313	22	ABG14394	Novel human diagno
40	38	48.1	328	22	AAV41174	Human polypeptide
41	38	48.1	414	22	AAV29180	Human PRO polypept
42	38	48.1	414	22	AAV87576	Human PRO3579. Ho
43	38	48.1	571	22	ABG27421	Novel human diagno
44	37.5	47.5	129	22	ABG02129	Novel human diagno
45	37	46.8	69	22	AAV39255	Propionibacterium

#### ALIGNMENTS

RESULT 1	
AAV30353	standard; Peptide; 15 AA.
XX	
XX	AAV30353;
XX	09-NOV-1999 (first entry)
XX	
DE	Epitope derived from pneumococcal surface adhesion A protein.
XX	
KW	Pneumococcal surface adhesion A protein; Psaa; monoclonal antibody;
KW	vaccine; Streptococcus pneumoniae infection.
XX	
OS	Streptococcus pneumoniae.
XX	
XX	WO9945121-A1.
XX	
PD	10-SEP-1999.
XX	
PF	26-FEB-1999; 99WO-US04326.
XX	
PR	02-MAR-1998; 98US-0076565.
XX	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;
PI	Zeller JL;
XX	
DR	WPI; 1999-540849/45.
XX	
PT	New peptides corresponding to Streptococcus pneumoniae Psaa, used
PT	for treating or preventing Streptococcus pneumoniae infection in a
PT	subject
XX	

```

PS Claim 6; Page 43; 58pp; English.
CC AAY30351-54 represent immunogenic peptides which are derived from
CC a pneumococcal surface adhesion A protein (PsaA). The specification
CC describes monoclonal antibodies which bind epitopes of the PsaA protein
CC (e.g present sequence). The peptides can be used in vaccines to prevent
CC Streptococcus pneumoniae infections. The antibodies of the invention
CC can also be used to detect S. pneumoniae in a sample or individual.
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 79; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 LVRRFVHRRPHVESQ 15
   |||||
Db 1 lvrrfvhrphvesq 15

RESULT 2
ID AAY30354 standard; Peptide: 15 AA.
XX
AC AAY30354;
XX
DT 09-NOV-1999 (first entry)
XX
DE Epitope derived from pneumococcal surface adhesion A protein.
XX
KM Pneumococcal surface adhesion A protein; PsaA; monoclonal antibody;
XX vaccine; Streptococcus pneumoniae infection.
XX
OS Streptococcus pneumoniae.
XX
PN M09945121-A1.
XX
PD 10-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04326.
XX
PR 02-MAR-1998; 98US-0076565.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
PI Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;
PI Zeller JL;
DR WPI: 1999-540849/45.
XX
XX New peptides corresponding to Streptococcus pneumoniae PsaA, used
XX for treating or preventing Streptococcus pneumoniae infection in a
XX subject
XX
PS Claim 6; Page 43; 58pp; English.
CC AAY30351-54 represent immunogenic peptides which are derived from
CC a pneumococcal surface adhesion A protein (PsaA). The specification
CC describes monoclonal antibodies which bind epitopes of the PsaA protein
CC (e.g present sequence). The peptides can be used in vaccines to prevent
CC Streptococcus pneumoniae infections. The antibodies of the invention
CC can also be used to detect S. pneumoniae in a sample or individual.
XX
SQ Sequence 15 AA;

Query Match 93.7%; Score 74; DB 20; Length 15;
Best Local Similarity 93.3%; Pred. No. 1.5e-06;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0
QY 1 LVRRFVHRRPHVESQ 15
   |||||

```

Db	1	lvrrfvhprhvesq	15
RESULT	3		
ID	AAB75292		
AC	AAB75292	standard; Protein; 156 AA.	
XX	AAB75292;		
DT	03-APR-2001	(first entry)	
DE	Gene 6 human secreted protein homologous amino acid sequence #11.		
XX	Human; immunosuppressive; antiarthritic; antirheumatic; nootropic;		
KW	antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;		
KW	neuroprotective; antibacterial; virucide; fungicide; opthalmological;		
KW	vulnerable; autoimmune disease; hyperproliferative disorder; cancer;		
KW	cardiovascular disorder; cerebrovascular disorder; infection;		
KW	nervous system disorder; ocular disorder; chemotaxis; food additive;		
KX	secreted protein.		
OS	Homo sapiens.		
PN	W0200077021-A1.		
XX	21-DEC-2000.		
PF	01-JUN-2000; 2000MO-US15135.		
PR	11-JUN-1999; 9905-0138632.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	(ROSE/) ROSEN C A.		
XX	Rosen CA, Ruben SM, Komatsoulis GA;		
DR	WPI; 2001-071257/08.		
PT	Nucleic acid molecules encoding human secreted proteins, used in		
PT	preventing, treating or ameliorating a disorder, e.g. Alzheimer's and		
PT	Parkinson's diseases and cancers -		
PS	Disclosure; Page 18-19; 530pp; English.		
XX	This invention relates to polynucleotide sequences AAF63789 - AAF63836		
CC	which encode human secreted proteins AAB75260 - AAB75287. Included in the		
CC	invention are protein sequences AAB75288 - AAB75341 which are fragments		
CC	of the secreted proteins and amino acid sequences with which these		
CC	fragments share homology. Examples of the activities of the proteins and		
CC	polynucleotides and the activities of their agonists and antagonists		
CC	include, immunosuppressive; antiarthritic; antirheumatic;		
CC	antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;		
CC	nootropic; neuroprotective; antibacterial; virucide; fungicide;		
CC	opthalmological; and vulnerary activity. The protein and polynucleotide		
CC	sequences, their agonists and antagonists may be useful for treating,		
CC	preventing and diagnosing diseases and disorders such as autoimmune		
CC	diseases e.g. rheumatoid arthritis, hyperproliferative disorders		
CC	e.g. neoplasms of the breast or liver, cardiovascular disorders		
CC	e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,		
CC	angiogenesis, nervous system disorders e.g. Alzheimer's disease,		
CC	infections caused by bacteria, viruses and fungi and ocular disorders		
CC	e.g. corneal infection. The polypeptides can also be used to aid wound		
CC	healing and epithelial cell proliferation, to prevent skin aging due to		
CC	sunburn, to maintain organs before transplantation, for supporting cell		
CC	culture of primary tissues, to regenerate tissues and in chemotaxis. The		
CC	polypeptides can also be used as a food additive or preservative to		
CC	increase or decrease storage capabilities. Included in the invention are		
CC	oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used		
CC	in the identification and characterisation of the DNA and protein		
CC	sequences of the invention.		
XX			
XX			
Sequence	156 AA;		
50			





QY 2 VAREVHRRPHE 13  
| | | | |  
DB 103 vrfvfkphvd 114

## RESULT 6

ABG18986  
ID ABG18986 standard; Protein; 273 AA.

XX AC ABG18986;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #18977.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS83173.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -

XX PS Claim 20; SEQ ID NO 49345; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPD  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 273 AA;

Query Match 65.8%; Score 52; DB 22; Length 273;  
Best Local Similarity 75.0%; Pred. No. 0.28;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VAREVHRRPHE 13  
| | | | |  
DB 124 vrfvfkphvd 135

## RESULT 7

ABG10923  
ID ABG10923 standard; Protein; 280 AA.

XX AC ABG10923;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #10914.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS75110.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -

XX PS Claim 20; SEQ ID NO 41282; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPD  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 280 AA;

Query Match 65.8%; Score 52; DB 22; Length 280;  
Best Local Similarity 75.0%; Pred. No. 0.28;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VAREVHRRPHE 13

Db 124 vrfvfhkphvd 135

RESULT 8  
ID ABG10922 standard; Protein; 282 AA.  
XX  
AC ABG10922;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10913.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS75109.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20: SEQ ID No 41281; 103pp; English.  
XX  
XS The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 282 AA:

Query Match 65.8%; Score 52; DB 22; Length 282;  
Best Local Similarity 75.0%; Pred. No. 0.29;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VRRFVRRPHE 13  
II III:IIII:

Db 103 vrfvfhkphvd 114

RESULT 9  
ID AAY59757 standard; Protein; 351 AA.  
XX  
AC AAY59757;  
XX  
DT 18-JAN-2000 (first entry)  
XX  
DE Human normal ovarian tissue derived protein 34.  
XX  
KW Human; ovary; screening; ovarian cancer; treatment.  
XX  
OS Homo sapiens.  
XX  
PN DE19816395-A1.  
PD 07-OCT-1999.  
XX  
PF 03-APR-1998; 98DE-1016395.  
XX  
PR 03-APR-1998; 98DE-1016395.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
XX  
DR WPI: 1999-552352/47.  
DR N-PSDB; AA241237.  
XX  
PT Nucleic acid sequences potentially useful in diagnosis or therapy of  
PT ovarian cancer -  
XX  
PS Claim 23; Page 229; 274pp; German.  
XX  
XS This invention describes novel nucleic acid sequences that are highly  
CC expressed in normal ovary tissue. Artificial chromosomes and cosmid  
CC clones containing the sequences can be used as gene transfer vehicles.  
CC The sequences can be used to produce DNA fragments containing  
CC full-length genes. Host cells transformed with the sequences can be used  
CC to produce polypeptides or polypeptide fragments, which can be used to  
CC screen phage displays for polypeptides that bind to them, or as tools for  
CC identifying agents active against ovarian cancer, or to prepare  
CC medicaments for treating ovarian cancer. The cDNA sequences can be used  
CC to obtain genomic genes, their promoters, enhancers, silencers, exon  
CC structures, intron structures and their splice variants. AAY59724-Y59837  
CC represent protein fragments encoded by the cDNA sequences represented in  
CC AA241222-241324 which are derived from normal human ovarian tissue.  
XX  
SQ Sequence 351 AA:

Query Match 65.8%; Score 52; DB 20; Length 351;  
Best Local Similarity 75.0%; Pred. No. 0.36;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VRRFVRRPHE 13  
II III:IIII:

RESULT 10  
ID ABG18985 standard; Protein; 142 AA.  
XX  
AC ABG18985;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #18976.  
XX

Query Match 65.8%; Score 52; DB 22; Length 282;  
Best Local Similarity 75.0%; Pred. No. 0.29;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VRRFVRRPHE 13  
II III:IIII:

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HXSE-) HXSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS83172.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 49344; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 SQ Sequence 142 AA;  
 Query Match 60.8%; Score 48; DB 22; Length 142;  
 Best Local Similarity 75.0%; Pred. No. 0.7;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 VRRVRRRRPVE 13  
 II :IIII  
 Db 3 vrfvrrpghvd 14  
 RESULT 11  
 AAY34665  
 ID AAY34665 standard; Protein; 580 AA.  
 XX  
 AC AAY34665;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Chlamydia pneumoniae protein not found in C. trachomatis.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-1B01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffiths R;  
 XX  
 DR WPI: 1999-357842/30.  
 DR  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 PT  
 PS Page 670-671; Disclosure; 1912pp; English.  
 XX  
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.  
 CC  
 SQ Sequence 580 AA;  
 Query Match 57.0%; Score 45; DB 20; Length 580;  
 Best Local Similarity 80.0%; Pred. No. 11;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 VRRVRRRRPH 11  
 III :IIIIII  
 Db 570 vrrlvrrrrph 579  
 RESULT 12  
 ABB62356  
 ID ABB62356 standard; Protein; 798 AA.  
 XX  
 AC ABB62356;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 13860.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 PN  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.





Query Match 53.2%; Score 42; DB 22; Length 175;  
 Best Local Similarity 57.1%; Pred. No. 10;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 LVRRFVHRPPHVES 14  
 : : | | | | | | | | :  
 Db 87 mlrrfnhrpfvlet 100

RESULT 15  
 AAB78972  
 ID AAB78972 standard; Protein: 203 AA.  
 XX  
 AC AAB78972:  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE C. glutamicum SRT protein sequence SEQ ID NO:204.  
 XX  
 XX Corynebacterium glutamicum; stress; resistance; tolerance; SRT;  
 KW fine chemical production; organic acid; proteinogetic amino acid;  
 KW nonproteinogetic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW evolutionary study; environmental hazard; fermentation.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN W0200100804-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000MO-IB00922.  
 XX  
 PR 25-JUN-1999; 990S-0141031.  
 PR 01-JUL-1999; 99DE-1030429.  
 PR 01-JUL-1999; 990S-0142692.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 09-JUL-1999; 99DE-1032209.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032914.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 31-AUG-1999; 990S-0151214.  
 PR 31-AUG-1999; 99DE-1041382.  
 XX  
 PA (BAD1 ) BASF AG.  
 XX  
 PI Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G, Lee H;  
 PI Kim H;  
 XX  
 DR N-PSDB; AAF71085.  
 XX  
 WPI: 2001-061972/07.  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a stress,  
 PT tolerance or resistance protein, for production or modulation of  
 PT production of fine chemicals, such as, e.g. amino acids, lipids,  
 PT carbohydrates, or enzymes -  
 XX  
 PS Claim 20; Page 374; 526pp; English.  
 XX  
 CC AAF70984 to AAF71133 encode the Corynebacterium glutamicum stress,  
 CC resistance, and tolerance (SRT) proteins given in AAB78871 to AAB79020.  
 CC The C. glutamicum SRT genes (I) can be used in vectors (II) for  
 CC expression in host cells and production of fine chemicals, such as, an  
 CC organic acid, a proteinogetic or nonproteinogetic amino acid (preferred),  
 CC a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a  
 CC saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic  
 CC compound, a vitamin, a cofactor, a polyketide, or an enzyme. The fine  
 CC chemical production can be modulated. The presence of (I) or the SRT

CC proteins (III) encoded by them are used for diagnosing the presence  
 CC or activity of Corynebacterium diptheriae. (I), (II), (III) and host  
 CC cells containing them can be used to map the genomes of organisms related  
 CC to C. glutamicum, to identify and localise C. glutamicum sequences of  
 CC interest, in evolutionary studies, in determination of SRT protein  
 CC regions required for function, in modulating the SRT protein activity,  
 CC and in modulating the activity of an SRT pathway. (II) are used to permit  
 CC C. glutamicum to survive in an environment that is normally  
 CC environmentally or chemically hazardous to it. (I) and protein molecules  
 CC encoded by it increase the survival of C. glutamicum to chemical and  
 CC environmental hazards and provide a means for continued growth and  
 CC multiplication in large scale fermentative growth conditions. By  
 CC increasing the growth rate or maintaining a normal growth rate in poor or  
 CC toxic conditions, the yield, production and/or efficiency or production  
 CC of fine chemicals from a culture may be increased.

XX  
 SQ Sequence 203 AA;

Query Match 53.2%; Score 42; DB 22; Length 203;  
 Best Local Similarity 57.1%; Pred. No. 12;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 LVRRFVHRPPHVES 14  
 : : | | | | | | | | :  
 Db 115 mlrrfnhrpfvlet 128

Search completed: August 15, 2002, 11:38:27  
 Job time: 268 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 11:38:58 ; Search time 26.88 seconds  
(without alignments)  
13.630 Million cell updates/sec

Title: US-09-613-092a-7  
Perfect score: 79  
Sequence: 1 LVRFVHRPHVESQ 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	65.8	244	2	US-08-933-750C-23
2	52	65.8	244	4	US-09-234-613-23
3	41	51.9	466	4	US-08-868-435-31
4	41	51.9	466	2	US-08-744-231-31
5	37	46.8	426	2	US-08-852-743-2
6	37	46.8	426	2	US-09-211-930-4
7	37	46.8	426	3	US-09-340-993-4
8	37	46.8	426	3	US-09-185-370-2
9	37	46.8	426	4	US-09-152-406-4
10	37	46.8	426	4	US-09-468-442-4
11	37	46.8	647	1	US-07-894-212A-8
12	37	46.8	649	1	US-07-894-212A-2
13	37	46.8	650	1	US-07-893-928A-1
14	36	45.6	362	1	US-08-415-751-6
15	36	45.6	375	2	US-08-506-864A-1
16	36	45.6	375	2	US-08-851-968-1
17	35	44.3	53	4	US-08-866-545-5
18	35	44.3	169	3	US-08-630-172-11
19	35	44.3	169	4	US-09-375-419-11
20	35	44.3	180	4	US-08-913-159-13
21	35	44.3	213	3	US-08-971-158-4
22	35	44.3	260	4	US-09-006-353A-8
23	35	44.3	462	3	US-08-486-099-117
24	35	44.3	462	3	US-08-360-107A-127
25	35	44.3	462	3	US-08-484-223B-117
26	35	44.3	462	3	US-08-919-597-117
27	35	44.3	462	3	US-08-475-668A-117

28	35	44.3	462	3	US-08-485-551A-117	Sequence 117, App
29	35	44.3	462	3	US-08-471-913A-117	Sequence 117, App
30	35	44.3	462	4	US-08-485-264A-117	Sequence 117, App
31	35	44.3	462	4	US-08-474-349A-117	Sequence 117, App
32	35	44.3	583	4	US-09-311-311C-24	Sequence 24, App
33	35	44.3	2259	4	US-09-413-814-70	Sequence 70, App
34	35	44.3	2439	3	US-09-335-409-7	Sequence 7, App
35	35	44.3	2439	4	US-09-568-102-7	Sequence 7, App
36	35	44.3	2439	4	US-09-567-969-7	Sequence 7, App
37	35	44.3	2439	4	US-09-568-480-7	Sequence 7, App
38	35	44.3	2439	4	US-09-568-486-7	Sequence 7, App
39	35	44.3	2439	4	US-09-568-472-7	Sequence 7, App
40	34.5	43.7	282	2	US-08-997-080-134	Sequence 134, App
41	34.5	43.7	282	2	US-08-997-362-134	Sequence 134, App
42	34.5	43.7	282	4	US-09-095-855-134	Sequence 134, App
43	34.5	43.7	282	4	US-09-324-542-134	Sequence 134, App
44	34.5	43.7	297	2	US-08-997-080-177	Sequence 177, App
45	34.5	43.7	297	2	US-08-997-362-177	Sequence 177, App

## ALIGNMENTS

RESULT 1  
US-08-933-750C-23  
Sequence 23, Application US/08933750C  
Patent No. 5932442  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Bandman, Olga  
APPLICANT: Shah, Purvi  
APPLICANT: Au-Young, Janice  
APPLICANT: Yue, Henry  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933.750C  
FILING DATE: September 23, 1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 244 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SPLNNOT04

CLONE: 1561587  
US-08-933-750C-23

Query Match 65.8%; Score 52; DB 2; Length 244;  
Best Local Similarity 75.0%; Pred. No. 0.1;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VRRFVHRPHVE 13  
DB 104 VRRFVHRPHVD 115

RESULT 2  
US-09-234-613-23  
Sequence 23, Application US/09234613  
Patent No. 6132973

## GENERAL INFORMATION:

APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Bandman, Olga  
APPLICANT: Shah, Purvi  
APPLICANT: Au-Young, Janice  
APPLICANT: Yue, Henry  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/234,613  
FILING DATE:

## CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750  
FILING DATE: September 23, 1997  
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

## TELEX:

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: SPLNOT04  
CLONE: 1561587

US-09-234-613-23

Query Match 65.8%; Score 52; DB 4; Length 244;  
Best Local Similarity 75.0%; Pred. No. 0.1;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VRRFVHRPHVE 13  
DB 104 VRRFVHRPHVD 115

DB 104 VRRFVHRPHVD 115

## RESULT 3

US-08-868-435-31  
Sequence 31, Application US/08868435  
Patent No. 6291221

## GENERAL INFORMATION:

APPLICANT: Van Loon, Adolphus  
APPLICANT: Mitchell, David  
TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: United States of America  
ZIP: 07110

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/868,435  
FILING DATE:

## CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/744,231  
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Kass, Alan P.  
REGISTRATION NUMBER: 32142

REFERENCE/DOCKET NUMBER: Case Docket 9339  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-4205  
TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 466 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

FEATURE:

NAME/KEY: misc-feature  
LOCATION: 204

OTHER INFORMATION: /note="potential N-glycosylation site"

FEATURE:

NAME/KEY: misc-feature  
LOCATION: 269

OTHER INFORMATION: /note="potential N-glycosylation site"

FEATURE:

NAME/KEY: misc-feature  
LOCATION: 335

OTHER INFORMATION: /note="potential N-glycosylation site"

FEATURE:

NAME/KEY: misc-feature  
LOCATION: 372

OTHER INFORMATION: /note="potential N-glycosylation site"

US-08-868-435-31

Query Match 51.9%; Score 41; DB 4; Length 466;  
Best Local Similarity 57.1%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LVRRFVHRPHVES 14  
DB 104 LVRRFVHRPHVDS 115

Db 12 LVALVSRNPHVDS 25

```

RESULT 4
US-08-744-231-31
; Sequence 31, Application US/08744231
; Patent No. 6358722
; GENERAL INFORMATION:
; APPLICANT: Van Loon, Adolphus
; APPLICANT: Mitchell, David
; TITLE OF INVENTION: POLYPEPTIDES WITH PHITASE ACTIVITY
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,231
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/424,757
; FILING DATE: 18-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kass, Alan P
; REGISTRATION NUMBER: 32142
; REFERENCE/DOCKET NUMBER: Case Docket 9339
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-4205
; TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 204
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 269
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 335
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 348
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 372
; OTHER INFORMATION: /note="potential N-glycosylation site"
; US-08-744-231-31

```

Query Match 51.9%; Score 41; DB 4; Length 466;  
 Best Local Similarity 57.1%; Pred. No. 16;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Oy 1 LVRRVHRRPHVES 14  
 || : | | | | | | |

Db 12 LVALVSRNPHVDS 25

```

RESULT 5
US-08-852-743-2
; Sequence 2, Application US/08852743
; Patent No. 5830699
; GENERAL INFORMATION:
; APPLICANT: Force, Thomas
; APPLICANT: Kyriakis, John M.
; APPLICANT: Pombo, Cella M.
; APPLICANT: Bonventre, Joseph
; TITLE OF INVENTION: SOK-1 AND METHODS OF USE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,743
; FILING DATE: 7-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/016,774
; FILING DATE: 7-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00786/327001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200134
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 426 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; US-08-852-743-2

```

Query Match 46.8%; Score 37; DB 2; Length 426;  
 Best Local Similarity 53.8%; Pred. No. 71;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Oy 2 VRRFVHRRPHVES 14  
 || : | | | | | | |  
 Db 412 VQRFSHRNHLTS 424

```

RESULT 6
US-09-211-930-4
; Sequence 4, Application US/09211930
; Patent No. 5962265
; GENERAL INFORMATION:
; APPLICANT: Tyrell E. No. 5962265r1s
; APPLICANT: William Craig Moore
; APPLICANT: David Shay Silberstein
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION
; TITLE OF INVENTION: SERINE/THREONINE KINASE
; FILE REFERENCE: PHM.70296
; CURRENT APPLICATION NUMBER: US/09/211,930
; CURRENT FILING DATE: 1998-12-15

```

```

? EARLIER APPLICATION NUMBER: GB 9726851.0
? EARLIER FILING DATE: 1997-12-19
? NUMBER OF SEQ ID NOS: 14
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 4
? LENGTH: 426
? type: prt
? ORGANISM: Homo sapiens
? OS-09-211-930-4

```

Query Match	46.88;	Score 37;	DB 2;	Length 426;
Best Local Similarity	53.88;	Pred. No. 71;		
Matches	7;	Conservative	2;	Mismatches 4;
				Indels 0;
				Gaps 0;

```

OY      2 VRRFVHRPPVES 14
         | : | | | | : |
Db      412 VQRFSHNRNHLTS 424

```

RESULT 7  
US-09-340-993-4  
; Sequence 4, Application US/09340993  
; Patent No. 6,074,330

```

: GENERAL INFORMATION:
: APPLICANT: Tyrell E. No. 6034228:is
: APPLICANT: William Craig Moore
: APPLICANT: David Shay Silberstein
: TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE
: FILE REFERENCE: PHM.70236.NI
: CURRENT APPLICATION NUMBER: US/09/340,993

```

```
?
?
? EARLIER APPLICATION NUMBER: GB 9726851.0 & US 09/211,993.0
? EARLIER FILING DATE: 1997-12-19 & 1998-12-15
? NUMBER OF SEO ID NOS: 14
? SOFTWARE: FastSeq for Windows Version 3.0
? SEO ID NO 4
? LENGTH: 426
? TYPE: PRT
? ORGANISM: Homo sapiens
? OS-09-340-993-4
```

Query Match	46.8%;	Score 37;	DB 3;	Length 426;
Best Local Similarity	53.8%;	Pred. NO. 71;		
Matches	7; Conservative	2; Mismatches	4; Indels	0; Gaps

```
QY      2 VRREVHRRPHVES 14
        |:| | | |:|
Db     412 VQFSHNRNHLTS 424
```

RESULT 8  
US-09-185-370-2  
; Sequence 2, Application US/09185370  
; Patent No. 6093560

GENERAL INFORMATION:  
APPLICANT: Force, Thomas  
APPLICANT: Kyriakis, John M.  
APPLICANT: Pombo, Cella M.  
APPLICANT: Bonventre, Joseph  
TITLE OF INVENTION: SOK-1 AND METHODS OF USE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston

STATE: MA  
COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

COMPUTER: IBM Compatible

```

: OPERATING SYSTEM: Windows95
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/185, 370
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/852, 743
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Fraser, Janis K.
: REGISTRATION NUMBER: 34, 819
: REFERENCE/DOCKET NUMBER: 00786/327001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ. ID NO.: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 426 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
:
: US-09-185-370-2

```

Query Match	46.8%;	Score 37;	DB 3;	Length 426;
Best Local Similarity	53.8%;	Pred. NO. 71;		
Matches	7;	Conservative	2;	Mismatches
			4;	Indels
				0;
				Gaps
				0;

```
QY      2 VRRFVHRPHEVES 14
      ||:| | | | : |
DB     412 VQRFSHNRNHLTS 424
```

RESULT 9  
US-09-152-406-4  
; Sequence 4, Application US/09152406

Patent No. 6265560  
GENERAL INFORMATION:  
APPLICANT: William Craig Moore  
APPLICANT: Tyrell E. No. 6265560is  
APPLICANT: David Shay Silberstein  
TITLE OF INVENTION: HUMAN STE20-LIKE STRESS ACTIVATED  
TITLE OF INVENTION: SERINE/THREONINE KINASE  
FILE REFERENCE: PHM.70272

```

? CURRENT APPLICATION NUMBER: US/09/152,406-4
? CURRENT FILING DATE: 1998-09-14
? EARLIER APPLICATION NUMBER: 9719920.2
? EARLIER FILING DATE: 1997-09-19
? NUMBER OF SEQ ID NOS: 9
? SOFTWARE: FASTSEQ for Windows Version 3.0.
? SEQ ID NO 4
? LENGTH: 426
? TYPE: PRT
? ORGANISM: Homo Sapiens
? OS-09-152-406-4

```

Query Match	46.88;	Score 37;	DB 4;	Length 426;
Best Local Similarity	53.88;	Pred. No. 71;		
Matches	7;	Conservative	2;	Mismatches 4;
				Indels 0;
				Gaps 0;

```
QY      2 VRRFVHRPHVES 14
          | : | | | : |
Db     412 VQRFSHNRNHLTS 424
```

RESULT 10

US-09-468-442-4

; Sequence 4, Application US/09468442  
; Patent No. 6300098

GENERAL INFORMATION:  
APPLICANT: Tyrell E. No. 630009815  
APPLICANT: William Craig Moore  
APPLICANT: David Shay Silberstein  
TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE  
FILE REFERENCE: PM.70236.NI  
CURRENT APPLICATION NUMBER: US/09/468,442  
CURRENT FILING DATE: 1999-12-21  
EARLIER APPLICATION NUMBER: US 09/340,993  
EARLIER FILING DATE: 1999-06-25  
EARLIER APPLICATION NUMBER: GB 9726851.0  
EARLIER FILING DATE: 1997-12-19  
EARLIER APPLICATION NUMBER: US 09/211,930  
EARLIER FILING DATE: 1998-12-15  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 426  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-468-442-4

Query Match 46.8%; Score 37; DB 4; Length 426;  
Best Local Similarity 53.8%; Pred. No. 71;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 VRRFVRRPHVES 14  
Db 412 VRRFVRRPHVES 424

RESULT 11  
US-07-894-212A-8  
Sequence 8, Application US/07894212A  
Patent No. 5366883  
GENERAL INFORMATION:  
APPLICANT: ASADA, KIYOZO  
APPLICANT: UEMORI, TAKASHI  
APPLICANT: MUKAI, HIROYUKI  
APPLICANT: KATO, IKUNOSHIN  
APPLICANT: LADERMAN, KENNETH  
APPLICANT: ANFINSEN, CHRISTIAN  
TITLE OF INVENTION: THE ALPHA-AMYLASE GENE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DABRY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON, D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/894,212A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16773  
REFERENCE/DOCKET NUMBER: 95469/C-1195  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
CUSH  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 647 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-894-212A-8

Query Match 46.8%; Score 37; DB 1; Length 647;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LVRRFVRRPHVESQ 15  
Db 435 LVRRFVRRPHVESQ 449

RESULT 12  
US-07-894-212A-2  
Sequence 2, Application US/07894212A  
Patent No. 5366883  
GENERAL INFORMATION:  
APPLICANT: ASADA, KIYOZO  
APPLICANT: UEMORI, TAKASHI  
APPLICANT: MUKAI, HIROYUKI  
APPLICANT: KATO, IKUNOSHIN  
APPLICANT: LADERMAN, KENNETH  
APPLICANT: ANFINSEN, CHRISTIAN  
TITLE OF INVENTION: THE ALPHA-AMYLASE GENE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DABRY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON, D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/894,212A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16773  
REFERENCE/DOCKET NUMBER: 95469/C-1195  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
CUSH  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 649 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-894-212A-2

Query Match 46.8%; Score 37; DB 1; Length 649;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LVRRFVRRPHVESQ 15  
Db 437 LVRRFVRRPHVESQ 451

RESULT 13  
US-07-893-928A-1  
Sequence 1, Application US/07893928A  
Patent No. 5578479

GENERAL INFORMATION:  
APPLICANT: LADERMAN, KENNETH  
APPLICANT: ANFINSEN, CHRISTIAN  
TITLE OF INVENTION: a-AMYLASE FROM HYPERTHERMOPHILIC  
TITLE OF INVENTION: ARCHAEABACTERIUM  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON, D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Tape  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/893,928A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL, N.  
REGISTRATION NUMBER: 16773  
REFERENCE/DOCKET NUMBER: 95470/C-1197  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 650 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-893-928A-1

Query Match 46.8%; Score 37; DB 1; Length 650;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LVRRFVHRPHVESQ 15  
DB 437 LARRMEHYHGYVESQ 451

RESULT 14  
US-08-415-751-6  
Sequence 6, Application US/08415751  
Patent No. 5643772  
GENERAL INFORMATION:  
APPLICANT: PETERSEN, CAROLYN  
APPLICANT: LEECH, JAMES  
APPLICANT: NELSON, RICHARD, C.  
APPLICANT: GUT, JIRI  
TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-  
TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA  
TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID  
TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND  
TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND  
TITLE OF INVENTION: DIAGNOSIS AND KIT  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PHILLIPS, MOORE, LEMPICO & FINLEY  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: California  
COUNTRY: United States of America  
ZIP: 94306-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage

COMPUTER: PC  
OPERATING SYSTEM: DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/415,751  
FILING DATE: 03-APR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/071,880  
FILING DATE: June 1, 1993  
APPLICATION NUMBER: 07/891,301  
FILING DATE: May 29, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hana Dolezalova  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480.19-2 (HND)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-1677  
TELEFAX: (415) 324-1678  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 362 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORGANISM: Cryptosporidium parvum  
FEATURE:  
NAME/KEY: Positions coded by nonsense codons are  
identified as Xaa.  
US-08-415-751-6

Query Match 45.6%; Score 36; DB 1; Length 362;  
Best Local Similarity 60.0%; Pred. No. 89;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 VRRFVHRPH 11  
DB 328 LREFVHRSH 337

RESULT 15  
US-08-506-864A-1  
Sequence 1, Application US/08506864A  
Patent No. 5834245  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: FUJIMURA, YOSHIYUKI  
TITLE OF INVENTION: PRITS PROTEINS AND DNA'S  
TITLE OF INVENTION: ENCODING THE SAME  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLYNN, THEIL, BOUTELL & TANIS, P.C.  
STREET: 2026 Rambling Road  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49008-1699  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: MS-DOS 5.0  
SOFTWARE: Wordperfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/506,864A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP6-178131  
FILING DATE: 29-JULY-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Terryence F. Chapman  
REGISTRATION NUMBER: 32549  
REFERENCE/DOCKET NUMBER: Furuya Case 1334  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (616) 381-1156  
TELEFAX: (616) 381-5465  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
LIBRARY: human fetal lung cDNA library  
US-08-506-864A-1

Query Match 45.6%; Score 36; DB 2; Length 375;  
Best Local Similarity 53.8%; Pred. NO. 93;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 3 RRFVRRPPHVESQ 15  
| | | : | | | |  
Db 229 RGFVYLQPHSEHQ 241

Search completed: August 15, 2002, 11:38:59  
Job time: 265 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:39:41 ; Search time 35.58 seconds  
(without alignments)  
40.510 Million cell updates/sec

Title: US-09-613-092a-7  
Perfect score: 79  
Sequence: 1 LVRRFVHRRPHVESQ 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	58.2	324	2 F69758	transcription regu
2	45	57.0	577	2 F72122	hypothetical prote
3	45	57.0	577	2 C86499	hypothetical prote
4	44	55.7	546	2 I49636	DNA-binding protei
5	43	54.4	72	2 T12874	hypothetical prote
6	41	51.9	282	2 G70482	panthothenate synth
7	41	51.9	465	2 A44498	radial spoke prote
8	41	51.9	481	2 T15372	hypothetical prote
9	40	50.6	138	2 H72534	hypothetical prote
10	40	50.6	1220	2 T42573	DNA-directed DNA p
11	40	50.6	1233	2 T30969	serine/threonine p
12	39	49.4	67	2 S14854	hypothetical prote
13	39	49.4	80	2 G81065	hypothetical prote
14	39	49.4	246	2 T16371	hypothetical prote
15	39	49.4	257	2 E70330	conserved hypochet
16	39	49.4	261	2 T51222	hypothetical prote
17	39	49.4	476	2 T40350	hypothetical prote
18	39	49.4	540	2 JC4916	signal transducing
19	39	49.4	548	2 JC4917	signal transducing
20	39	49.4	954	2 G71496	hypothetical prote
21	39	49.4	1207	2 H87475	urea amidolysase-re
22	39	49.4	1792	2 T13939	myosin V - fruit f
23	38	48.1	248	2 S30194	ribosomal protein
24	38	48.1	363	2 C86214	hypothetical prote
25	38	48.1	384	2 T45926	hypothetical prote
26	38	48.1	434	2 F75425	tRNA nucleotidyltr
27	38	48.1	661	2 S75005	sensory transducti
28	38	48.1	688	2 T21641	hypothetical prote
29	38	48.1	837	2 T19825	hypothetical prote

30	37.5	47.5	241	2 H82072	DNA repair protein
31	37	46.8	167	2 S76074	hypothetical prote
32	37	46.8	235	2 A72594	hypothetical prote
33	37	46.8	244	2 T43958	hypothetical prote
34	37	46.8	274	2 AE1619	hypothetical prote
35	37	46.8	274	2 A11256	hypothetical prote
36	37	46.8	292	2 AD0362	probable lysR-fam1
37	37	46.8	426	2 S71886	Ste20-like protein
38	37	46.8	589	2 T38086	serine/threonine-p
39	37	46.8	590	2 A10094	ABC-transporter tr
40	37	46.8	600	2 D83683	L-glutamine-D-fruc
41	37	46.8	600	2 T14613	hypothetical prote
42	37	46.8	649	2 A49512	alpha-amylase (EC
43	37	46.8	651	2 T03889	Na+/Ca2+,K+-exchan
44	37	46.8	675	1 FYFEP	brown protein - fr
45	37	46.8	1108	2 T17455	translation initia

## ALIGNMENTS

RESULT 1  
F69758  
transcription regulator lysR family homolog ycgK - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: F69758  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabrel, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Firtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gal  
lechl, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holst, M.; Kunita, K.; Lapidus, A.; Lardino  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Potte  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchl, M.; Tamakoshi, A.; Tanaka, T.; Terstape, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zundstein, E.; Yoshikawa, H.; Danolin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033  
A:Accession: F69758  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-324 <KUN>  
A:Cross-references: GB:299105; GB:AL009126; NID:q2632457; PIDN:CAB12111.1; PID:q26326  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ycgK  
C:Superfamily: probable transcription regulator lysR

Query Match 58.2% Score 46; DB 2; Length 324;  
Best Local Similarity 60.0%; Pred. No. 2.6;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LVRRFVHRRPHVESQ 15  
DB 109 LKQFVARYPHEVO 123

RESULT 2  
F72122  
hypothetical protein CP0708 [imported] - Chlamydomonas reinhardtii (strains CHL029 and  
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: F72122; D81548  
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.  
A:Reference number: A72000; MUID:99206606

A:Accession: F72122  
A:Molecule type: DNA  
A:Residues: 1-577 <ARN>  
A:Cross-references: GB:AE001592; GB:AE001363; NID:94376322; PIDN:AA018219.1; PID:9437632  
A:Experimental source: strain CMI029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of *Chlamydia trachomatis* Mopn and *Chlamydia pneumoniae* AR39.  
A:Reference number: AB1500; MUID:20150255  
A:Accession: DB1548  
A:Molecule type: DNA  
A:Residues: 1-577 <REA>  
A:Cross-references: GB:AE002229; GB:AE002161; NID:97189613; PIDN:AAF38516.1; PID:9718962  
A:Experimental source: strain AK39, HL cells  
C:Genetics:  
A:Gene: CP00066; CP0708

Query Match 57.0%; Score 45; DB 2; Length 577;  
Best Local Similarity 80.0%; Pred. No. 7;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VRRFVHRPH 11  
|||:|||||  
DB 567 VRRNLHRPH 576

RESULT 3  
C68499  
Hypothetical protein CPJ0066 [imported] - *Chlamydia pneumoniae* (strain J138)  
C:Species: *Chlamydia pneumoniae*, *Chlamydia pneumoniae*  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: C68499  
R:Shiral, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shida, T.; IS Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of *chlamydia pneumoniae* J138.  
A:Reference number: AB6491; MUID:20330349  
A:Accession: C68499  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-577 <STO>  
A:Cross-references: GB:BA000008; NID:98978440; PIDN:BA98277.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: CPJ0066

Query Match 57.0%; Score 45; DB 2; Length 577;  
Best Local Similarity 80.0%; Pred. No. 7;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VRRFVHRPH 11  
|||:|||||  
DB 567 VRRNLHRPH 576

RESULT 4  
I49636  
DNA-binding protein - mouse  
C:Species: *Mus musculus* (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 01-Dec-2000  
C:Accession: I49636  
R:Wick, M.J.; Ann, D.K.; Lee, N.M.; Loh, H.H.  
Gene 152, 227-232, 1995  
A:Title: Isolation of a cDNA encoding a novel zinc-finger protein from neuroblastoma x A:Reference number: I49636; MUID:95137393  
A:Accession: I49636  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-546 <RES>  
A:Cross-references: GB:U20450; NID:9437443; PIDN:AAA65196.1; PID:9437444  
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 55.7%; Score 44; DB 2; Length 546;  
Best Local Similarity 58.3%; Pred. No. 9.8;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 RVVHRPHVRSQ 15  
|:|||||:  
DB 117 RQIHRPHTEEK 128

RESULT 5  
T12874  
Hypothetical protein yqoN - *Bacillus subtilis* phage SPBC2  
C:Species: *Bacillus subtilis* phage SPBC2  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 15-Oct-1999  
C:Accession: T12874; F69920  
R:Jazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Marel, C.; Karamata, D. submitted to the EMBL Data Library, August 1997  
A:Description: The complete nucleotide sequence of the *Bacillus subtilis* SPBctac2 pro A:Reference number: Z17583  
A:Accession: T12874  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-72 <LAZ>  
A:Cross-references: EMBL:AF020713; NID:93025478; PID:93025588; PIDN:AA013083.1  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Galizzi, A.; Gal tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M Koeter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete Rieger, M.; Rivalta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl A:Authors: Schleich, S.; Schroeter, R.; Sciofione, F.; Sekiguchi, J.; Sekowska, A.; Se kouchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiya T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis* A:Reference number: A69580; MUID:98044033  
A:Accession: F69920  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-72 <KUN>  
A:Cross-references: GB:299114; GB:AL009126; NID:92634230; PIDN:CAB13949.1; PID:el1855  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yqoN  
A:Note: yqoN

Query Match 54.4%; Score 43; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 FVHRPH 11  
|||||||  
DB 60 FVHRPH 66

RESULT 6  
G70482  
pantothenate synthetase - *Aquifex aeolicus*  
C:Species: *Aquifex aeolicus*  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 29-Sep-1999  
C:Accession: G70482  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V. Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. A:Reference number: A70300; MUID:98196666

A:Accession: G70482  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-282 <NOF>  
A:Cross-references: GB:AE000774; NID:g2984324; PIDN:AAC07847.1; PID:g2984326; GB:AE00065  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: panc  
C:Superfamily: pantoate-beta-alanine ligase

Query Match 51.9%; Score 41; DB 2; Length 282;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VRRFVHRRPHE 13  
:|:|:|:|:|:  
Db 228 IRAFIERHPHV 239

RESULT 7  
A44498  
radial spoke protein 4 - Chlamydomonas reinhardtii  
C:Species: Chlamydomonas reinhardtii  
C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C:Accession: A44498  
R:Curry, A.M.; Williams, B.D.; Rosenbaum, J.L.  
Mol. Cell. Biol. 12, 3967-3977, 1992  
A:Title: Sequence analysis reveals homology between two proteins of the flagellar radial  
A:Reference number: A44498; MUID:92375065  
A:Accession: A44498  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-465 <CUR>  
A:Cross-references: GB:M87526; NID:g167435; PIDN:AAA33092.1; PID:g167436  
A:Note: Sequence extracted from NCBI backbone (NCBIN:111696, NCBIPI:111697)

Query Match 51.9%; Score 41; DB 2; Length 465;  
Best Local Similarity 63.6%; Pred. No. 27;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 FVHRPHEVSO 15  
:|:|:|:|:|:  
Db 319 WNVHVRPHLSQ 329

RESULT 8  
T15372  
hypothetical protein C01F1.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15372  
R:Johnson, D.  
submitted to the EMBL Data Library, May 1996  
A:Description: The sequence of C. elegans cosmid C01F1.  
A:Reference number: Z18338  
A:Accession: T15372  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-481 <OH>  
A:Cross-references: EMBL:U58761; NID:g1330391; PID:g1330397; PIDN:AAB00717.1; GSPDB:GN00  
A:Experimental source: strain Bristol N2; clone C01F1  
C:Genetics:  
A:Gene: CESP.C01F1.1  
A:Map position: 2  
A:Insertions: 263/1; 287/2; 428/3

Query Match 51.9%; Score 41; DB 2; Length 481;  
Best Local Similarity 42.9%; Pred. No. 28;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VRRFVHRRPHEVSO 15  
|:|:|:|:|:|:  
Db 416 VRRYLRKRPHTTKE 429

RESULT 9  
H72534  
hypothetical protein APE1565 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: H72534  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jin-no, K.; Tawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
A:Reference number: A72450; MUID:99310339  
A:Accession: H72534  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-138 <KAW>  
A:Cross-references: DDBJ:AF000062; NID:g5105244; PIDN:BAAB0565.1; PID:d1044351; PID:g  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1565

Query Match 50.6%; Score 40; DB 2; Length 138;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 RRVVHRRPH 11  
|:|:|:|:|:|:  
Db 66 RSVVHRQPH 74

RESULT 10  
T42573  
DNA-directed DNA polymerase (EC 2.7.7.7) - equine herpesvirus 4 (strain NS80567)  
C:Species: equine herpesvirus 4  
A:Variety: strain NS80567  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T42573  
R:Teliford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.  
J. Gen. Virol. 79, 1197-1203, 1998  
A:Title: The DNA sequence of equine herpesvirus-4.  
A:Reference number: Z22173; MUID:98264497  
A:Accession: T42573  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1220 <TEL>  
A:Cross-references: EMBL:AF030027; NID:g2605950; PIDN:AAC59546.1; PID:g2605974  
A:Experimental source: strain NS80567  
C:Genetics:  
A:Gene: 30  
C:Superfamily: herpesvirus DNA-directed DNA polymerase  
C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 50.6%; Score 40; DB 2; Length 1220;  
Best Local Similarity 50.0%; Pred. No. 11e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LVRRFVHRRPHEV 14  
|:|:|:|:|:|:  
Db 1165 LKRFPERPHKTS 1178

RESULT 11  
T30989  
serine/threonine protein kinase NIK - mouse  
N:Alternate names: Nck interacting kinase  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30989  
 R:Su, Y.C.; Han, J.; Xu, S.; Cobb, M.; Skolnik, E.Y.  
 EMBL J. 16, 1279-1290, 1997  
 A:Title: NIK is a new Ste20-related kinase that binds NCK and MEK1 and activates the SA  
 A:Reference number: Z20954; MUID:97280817  
 A:Accession: T30989  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1233 <SOY>  
 A:Cross-references: EMBL:088984; NID:g1872545; PID:g1872546; PIDN:AAC53165.1  
 C:Keywords: protein kinase

Query Match 50.6%; Score 40; DB 2; Length 1233;  
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 HRRPVHESQ 15  
 ||||| : |  
 DB 496 HRRPVHQQ 504

RESULT 12  
 S14854

hypothetical protein 1 - yeast (*Pichia angusta*)  
 C:Species: *Pichia angusta*  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Apr-2000  
 C:Accession: S14854  
 R:Krutilla, A.I.; Seregina, S.A.; Tikhomirova, L.P.; Kryukov, V.M.  
 submitted to the EMBL Data Library, April 1991  
 A:Description: Nucleotide sequence of *Hansenula polymorpha* DNA region complementing DAK.  
 A:Reference number: S14854  
 A:Accession: S14854  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-67 <RRU>  
 A:Cross-references: EMBL:X58862; NID:g2762; PIDN:CAA1665.1; PID:g2763

Query Match 49.4%; Score 39; DB 2; Length 67;  
 Best Local Similarity 77.8%; Pred. No. 8.2;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 HRRPVHESQ 15  
 ||||| : |  
 DB 12 HRRPVHQA 20

RESULT 13

hypothetical protein NMB1589 [imported] - *Neisseria meningitidis* (strain MC58 serogroup  
 C:Species: *Neisseria meningitidis*  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: G81065  
 R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 Li, H.; Qin, H.; Yamachyan, J.; Gill, J.; Scarlato, V.; Masiagnani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A:Reference number: A81000; MUID:20175755  
 A:Accession: G81065  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-80 <RET>  
 A:Cross-references: GB:AE002509; GB:AE002098; NID:g7226832; PIDN:AAFA1942.1; PID:g722683  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB1589

Query Match 49.4%; Score 39; DB 2; Length 80;  
 Best Local Similarity 70.0%; Pred. No. 9.8;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 VRRFVHRRPH 11  
 ||:||||| :  
 DB 62 VRRFVHRRQY 71

RESULT 14  
 T16371

hypothetical protein F45E12.1 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16371  
 R:Latreille, P.  
 submitted to the EMBL Data Library, June 1995  
 A:Description: The sequence of *C. elegans* cosmid F45E12.  
 A:Reference number: Z18501  
 A:Accession: T16371  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-246 <LAT>  
 A:Cross-references: EMBL:U29536; NID:g868261; PID:g868266; PIDN:AAA68794.1; CESP:F45E  
 C:Genetics:  
 A:Gene: CESP:F45E12.1  
 A:Introns: 29/3; 53/1; 92/3; 120/2; 151/2; 193/1

Query Match 49.4%; Score 39; DB 2; Length 246;  
 Best Local Similarity 58.3%; Pred. No. 31;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 VRRFVHRRPHE 13  
 | : ||||| :  
 DB 100 VRRFVHRRPHVD 111

RESULT 15

E70330  
 conserved hypothetical protein aq\_342 - *Aquifex aeolicus*  
 C:Species: *Aquifex aeolicus*  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 20-Aug-1999  
 C:Accession: E70330  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
 V.  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.  
 A:Reference number: A70300; MUID:98196666  
 A:Accession: E70330  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-257 <AOE>  
 A:Cross-references: GB:AE00684; NID:g2983009; PIDN:AAC06636.1; PID:g2983025; GB:AE00  
 A:Experimental source: strain VFS  
 C:Genetics:  
 A:Gene: aq\_342  
 C:Superfamily: beta-lactamase regulatory protein; beta-lactamase regulatory protein h  
 F.17-248/Domain: beta-lactamase regulatory protein homology <BLR>

Query Match 49.4%; Score 39; DB 2; Length 257;  
 Best Local Similarity 50.0%; Pred. No. 33;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LVRRFVHRRPHV 12  
 ||: : |||||  
 DB 89 LVKKLIERHRRHV 100

Search completed: August 15, 2002, 11:39:42  
 Job time: 203 sec



**THIS PAGE BLANK (USPTO)**



```

OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -I- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +
CC diphosphate + (R)-pantothenate.
CC -I- PATHWAY: Pantothenate biosynthesis; last step.
CC -I- SIMILARITY: BELONGS TO THE PANTOTHENATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000774; AAC07847.1; -.
DR InterPro; IPR003721; Pantoate_ligase.
DR Pfam; PF02569; Pantoate_ligase; 1.
KW Pantothenate biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 282 AA; 32433 MW; 7E71B941C158DCB CRC64;

Query Match 51.9%; Score 41; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 5.9;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VRFVRRRHVE 13
:| | | | |
DB 228 IRAFIERHPVK 239

RESULT 3
RSP4_CHLRE STANDARD; PRT; 465 AA.
ID RSP4_CHLRE
AC Q01656;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Flagellar radial spoke protein 4.
GN RSP4.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21GR;
RA MEDLINE=92375065; PubMed=1508197;
RA Curry A.M., Williams B.D., Rosenbaum J.L.;
RT "Sequence analysis reveals homology between two proteins of the
RT flagellar radial spoke."
RL Mol. Cell. Biol. 12:3967-3977(1992).
CC -I- FUNCTION: FLAGELLAR RADIAL SPOKES CONTRIBUTE TO THE REGULATION
CC OF DYNEIN ARM ACTIVITY AND THUS THE PATTERN OF FLAGELLAR BENDING.
CC THEY CONSIST OF A THIN STALK, WHICH IS ATTACHED TO THE A SUBTIBER
CC OF THE OUTER DOUBLET MICROTUBULE, AND A BULBOUS HEAD, WHICH IS
CC ATTACHED TO THE STALK AND APPEARS TO INTERACT WITH THE
CC PROTECTIONS FROM THE CENTRAL PAIR OF MICROTUBULES.
CC -I- SUBUNIT: THE RADIAL SPOKE HEAD IS MADE OF FIVE DIFFERENT
CC POLYPEPTIDES (RSP1, RSP4, RSP6, RSP9, AND RSP10).
CC -I- SUBCELLULAR LOCATION: RADIAL SPOKE.
CC -I- SIMILARITY: TO THE FLAGELLAR RADIAL SPOKE PROTEIN 6.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M87526; AAA33092.1; -.
DR PIR; A44498; A44498.
KW Flagella.
SQ SEQUENCE 465 AA; 49798 MW; A23AFB030CDB3E29 CRC64;

Query Match 51.9%; Score 41; DB 1; Length 465;
Best Local Similarity 63.6%; Pred. No. 10;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 FVHRPHVESQ 15
:| | | | |
DB 319 WVHRPHLKSQ 329

RESULT 4
RS6_DROME STANDARD; PRT; 248 AA.
ID RS6_DROME
AC P29327; Q94993; Q9W3N4;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 40S ribosomal protein S6.
GN RPS6 OR L(1)AIR8 OR HEN OR CG10944.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93087515; PubMed=1454811;
RA Watson K.L., Konrad K.D., Woods D.F., Bryant P.J.;
RT "Drosophila homolog of the human S6 ribosomal protein is required for
RT tumor suppression in the hematopoietic system."
RL Proc. Natl. Acad. Sci. U.S.A. 89:11302-11306(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93192329; PubMed=8448211;
RA Spencer T.A., Mackie G.A.;
RT "The nucleotide sequence of a cloned cDNA encoding ribosomal protein
RT S6 from Drosophila melanogaster."
RL Biochim. Biophys. Acta 1172:332-334(1993).
RN [3]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=94018587; PubMed=8412647;
RA Stewart M.J., Denell R.;
RT "The Drosophila ribosomal protein S6 gene includes a 3' triplication
RT that arose by unequal crossing-over."
RL Mol. Biol. Evol. 10:1041-1047(1993).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```



RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., Moleod M.P., Mobergson D.,  
RA Merklov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN CONTROLLING CELL GROWTH  
CC AND PROLIFERATION THROUGH THE SELECTIVE TRANSLATION OF PARTICULAR  
CC CLASSES OF MRNA.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B/C; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- PTM: RIBOSOMAL PROTEIN S6 IS THE MAJOR SUBSTRATE OF PROTEIN  
CC KINASES IN EUKARYOTE RIBOSOMES. THE PHOSPHORYLATION IS STIMULATED  
CC BY GROWTH FACTORS, TUMOR PROMOTING AGENTS, AND MITOGENS. IT IS  
CC DEPHOSPHORYLATED AT GROWTH ARREST.  
CC -1- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L01658; AAC34306.1; -;  
DR EMBL: L07881; AAA28871.1; -;  
DR EMBL: L02074; AAB05982.1; -;  
DR EMBL: L02074; AAB05983.1; -;  
DR EMBL: L02074; AAB05984.1; -;  
DR EMBL: L02075; AAB05985.1; -;  
DR EMBL: AEO03442; AAF46288.1; ALT\_SEQ.  
DR PIR: A47240; A47240.  
DR PIR: S30194; S30194.  
DR EMBL: F0000004922; RPS6.  
DR InterPro: IPR001377; Ribosomal\_S6E.  
DR Pfam: PF01092; Ribosomal\_S6e; 1.  
DR ProDom: PD003460; Ribosomal\_S6e; 1.  
DR PROSITE: PS00578; RIBOSOMAL\_S6e; 1.  
DR Ribosomal protein; Phosphorylation.  
KW VARSPLIC 73 200  
FT VRLKKGSGCYPRPRGEGKRRKSRVCIIDANNVYALVY  
FT LKKGKNDIPGLTDTTIPRRIGPRKASIRKLYNLSKEDVVR  
FT RFVVRRLPAKDKKAKSPKIORLTLPVIOKRRRIAL  
FT KKKRQ -> LRLKTHSCHPRCNKRYKCKTYAKTIVEAN  
FT VSAVLTVLKNPSPLGVRSSNLSIKTYLLEDEDEVIP  
FT VKLQRRHQRKRONATKALAEVYKLVLRKREKSKANGRYV  
FT TIRKPKSSVSGKK (IN ISOFORM B).  
FT MISSING (IN ISOFORM B).  
SQ VARSPLIC 201 248  
SQ SEQUENCE 248 AA; 28407 MW; 4E78127C33B3BD CRC64;

Query Match 48.1%; Score 38; DB 1; Length 248;  
Best Local Similarity 88.9%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VRRFVRRP 10  
Db 153 VRRFVRRP 161  
RESULT 5  
04F3\_HUMAN  
ID 04F3\_HUMAN STANDARD; PRT; 312 AA.  
AC 095013;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Olfactory receptor 4F3.  
GN OR4F3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Courtney L., Langston Y., Stoneking T., Drone K., Shih K.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PUTATIVE ODOURANT RECEPTOR.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AC004908; AAD05195.1; -;  
DR InterPro: IPR000276; GPCR\_Rhodopsn.  
DR Pfam: PF00001; 7tm.1; 1.  
DR PRINTS: PR00237; GPCR\_RHODOPSIN.  
DR PROSITE: PS00237; G\_PROTEIN\_RECPT\_F1\_1;  
DR PROSITE: PS50262; G\_PROTEIN\_RECPT\_F1\_2; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
KW Multigene family; Olfaction.  
FT DOMAIN 1 25  
FT TRANSSEM 26 49  
FT DOMAIN 50 57  
FT TRANSSEM 58 79  
FT DOMAIN 80 100  
FT TRANSSEM 101 120  
FT DOMAIN 121 139  
FT TRANSSEM 140 158  
FT DOMAIN 159 195  
FT TRANSSEM 196 219  
FT DOMAIN 220 235  
FT TRANSSEM 236 258  
FT DOMAIN 259 269  
FT TRANSSEM 270 289  
FT DOMAIN 290 312  
FT DISULFID 97 189  
FT CARBOHYD 5 5  
SQ SEQUENCE 312 AA; 35087 MW; ABEE3F4ABA43F86 CRC64;  
Query Match 48.1%; Score 38; DB 1; Length 312;  
Best Local Similarity 63.6%; Pred. No. 21;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 5 FVYRPHVNSQ 15  
Db 256 FVYRPHVNSQ 266

```

RESULT 6
ID CGB3_CAEEL STANDARD; PRT; 385 AA.
AC 010654;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE G2/mitotic-specific cyclin B3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabditoidea;
OC Rhabdillidae; Peloderinae; Caenorhabditis.
NCBI_Taxid=6239;
RN NCB1
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-95403581; PubMed=7545687;
RA Kreutzer M.A., Richards J.P., de Silva-Udawatla M.N.,
RA Temenak J.J., Knoblich J.A., Lehner C.F., Bennett K.L.;
RT "Caenorhabditis elegans cyclin A- and B-type genes: a cyclin A
RT multigene family, an ancestral cyclin B3 and differential germline
RT expression.";
RT J. Cell Sci. 108:2415-2424(1995).
CC -!- FUNCTION: COULD BE INVOLVED AT THE G2/M (MITOSIS) TRANSITION.
CC INTERACTS WITH THE CDC2 AND CDK2 PROTEIN KINASES. G2/M CYCLINS
CC ACCUMULATE STEADILY DURING G2 AND ARE ABRUPTLY DESTROYED AT
CC MITOSIS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, U21282; AAA84395.1; -.
CC InterPro: IPR000553; Cyclin.
CC Pfam: PF00134; cyclin_1.
CC SMART: SM00385; cyclin_C_1.
CC SMART: PS00292; CYCLINS; FALSE_NEG.
CC PROSITE: PS00292; CYCLINS; Mitosis; Nuclear protein.
CC Cyclin; Cell cycle; Cell division; Mitosis; Nuclear protein.
CC SEQUENCE 385 AA; 44757 MW; 74CEDA847A76B695 CRC64;
SQ

```

```

Query Match 48.1%; Score 38; DB 1; Length 385;
Best Local Similarity 41.7%; Pred. No. 27;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 2 VRFVHRPHEVE 13
|:::|:|:|
DB 129 VAKYLKHPHEVD 140

```

```

RESULT 7
ID YLE5_CAEEL STANDARD; PRT; 837 AA.
AC P46941;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 94.2 kDa protein C38D4.5 in chromosome III.
GN C38D4.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabditoidea;
OC Rhabdillidae; Peloderinae; Caenorhabditis.
NCBI_Taxid=6239;
RN NCB1
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-BRISTOL N2;
RA Coles L.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: C-TERMINAL TO CHIMAERIN.
CC -!- SIMILARITY: CONTAINS 1 MW DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, Z46241; CAA86318.1; -.
CC Wormpep: C38D4.5; CE00918.
CC InterPro: IPR001849; PH.
CC InterPro: IPR00198; RhogAP.
CC InterPro: IPR001202; WW.
CC Pfam: PF00169; PH; 1.
CC Pfam: PF00620; RhogAP; 1.
CC Pfam: PF00397; WW; 1.
CC SMART: SM00233; PH; 1.
CC SMART: SM00324; RhogAP; 1.
CC SMART: SM00456; WW; 1.
CC PROSITE: PS01159; WW_DOMAIN_1; 1.
CC PROSITE: PS50020; WW_DOMAIN_2; 1.
CC PROSITE: PS50003; PH_DOMAIN_1.
CC Hypothetical protein.
CC DOMAIN 36 129
CC FT DOMAIN 36 129
CC PH 386 505
CC SEQUENCE 837 AA; 94196 MW; D1895E622D1F5997 CRC64;
SQ

```

```

Query Match 48.1%; Score 38; DB 1; Length 837;
Best Local Similarity 61.5%; Pred. No. 60;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 2 VRFVHRPHEVES 14
|:::|:|:|
DB 583 LRRFFRTPVES 595

```

```

RESULT 8
ID VCOM_ADECC STANDARD; PRT; 421 AA.
AC 065952;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Minor core protein (protein V).
GN PV.
OS Canine adenovirus type 1 (strain CL1).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
NCBI_Taxid=69150;
RN NCB1
RP Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
[1]
[1] SEQUENCE FROM N.A.
RA Campbell J.B., Zhao Y.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, U55001; AAB05440.1; -.
CC Core protein; Late protein.
CC SEQUENCE 421 AA; 47538 MW; C84536291CDDDEF9 CRC64;
SQ

```

Query Match 46.8%; Score 37; DB 1; Length 421;  
 Best Local Similarity 70.0%; Pred. No. 43;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRRFVHRRPH 11  
 |||| |||  
 Db 66 VRRFAPRRPY 75

RESULT 9  
 VCOM\_ADECR STANDARD; PRT; 421 AA.  
 AC 096685;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Minor core protein (Protein V).  
 GN PV.  
 OS Canine adenovirus type 1 (strain R1261).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=69151;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97275900; PubMed=9129661;  
 RA Morrison M.D., Onions D.E., Nicolson L.;  
 RT "Complete DNA sequence of canine adenovirus type 1.";  
 RL J. Gen. Virol. 78:873-878(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Y07760; CAA69063.1; -  
 KM Core protein: Late protein.  
 SQ SEQUENCE 421 AA; 47649 MW; A5F53E18748211F CRC64;

Query Match 46.8%; Score 37; DB 1; Length 421;  
 Best Local Similarity 70.0%; Pred. No. 43;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRRFVHRRPH 11  
 |||| |||  
 Db 66 VRRFAPRRPY 75

RESULT 10  
 ST25\_HUMAN STANDARD; PRT; 426 AA.  
 AC 000506; Q15522;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Serine/threonine protein kinase 25 (EC 2.7.1.37) (Sterile 20/oxidant  
 DE stress-response kinase 1) (Ste20/oxidant stress response kinase-1)  
 DE (SOK-1) (Ste20-like kinase).  
 GN STK25 OR SOK1 OR YSK1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97042345; PubMed=887545;  
 RA Pombo C.M., Bonventre J.V., Molnar A., Kyriakis J., Force T.;  
 RT "Activation of a human Ste20-like kinase by oxidant stress defines a  
 RT novel stress response pathway".  
 RL EMBO J. 15:4537-4546(1996).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97304522; PubMed=9160885;  
 RA Osada S.-I., Izawa M., Saito R., Mizuno K., Suzuki A., Hirai S.-I.,  
 RA Ohno S.;  
 RT "YSK1, a novel mammalian protein kinase structurally related to Ste20  
 RT and SPS1, but is not involved in the known MAPK pathways.";  
 RL Oncogene 14:2047-2057(1997).  
 CC -1- FUNCTION: OXIDANT STRESS-ACTIVATED SERINE/THREONINE KINASE THAT  
 CC MAY PLAY A ROLE IN THE RESPONSE TO ENVIRONMENTAL STRESS.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -1- ENZYME REGULATION: ACTIVATED BY PHOSPHORYLATION, PROBABLY  
 CC AUTOPHOSPHORYLATION. THE C-TERMINAL NON-CATALYTIC REGION INHIBITS  
 CC THE KINASE ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: CYTOSOL;PLASMC.  
 CC -1- TISSUE SPECIFICITY: UBIGUOUSLY EXPRESSED. HIGHEST LEVELS ARE  
 CC FOUND IN TESTIS, LARGE INTESTINE, BRAIN AND STOMACH FOLLOWED BY  
 CC HEART AND LUNG.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC STE20 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X99325; CAA67700.1; -  
 DR EMBL: D63780; BAA20420.1; -  
 DR HSSP: 000534; 1B17.  
 DR MIM: 602255; -  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002290; Ser\_Thr\_Pkinase.  
 DR Pfam: PF00069; Pkinase.1.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_SF; FALSE\_NEG.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR KIM: K00001; Serine/threonine-protein kinase; ATP-binding;  
 KW phosphorylation.  
 FT DOMAIN 20 270  
 FT NP\_BIND 26 34 ATP (BY SIMILARITY).  
 FT BINDING 49 49 ATP (BY SIMILARITY).  
 FT ACT\_SITE 140 140 BY SIMILARITY.  
 FT CONFLICT 347 348 EP -> DA (IN REF. 1).  
 SQ SEQUENCE 426 AA; 48111 MW; 183CE5700FCEA716 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 426;  
 Best Local Similarity 53.8%; Pred. No. 44;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VRRFVHRRPH 14  
 |||| | | | |  
 Db 412 VQRFSHNRNHLTS 424

RESULT 11  
 SHK2\_SCHPO STANDARD; PRT; 589 AA.  
 ID SHK2\_SCHPO  
 AC 010056;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Serine/threonine-protein kinase shk2 (EC 2.7.1.1).  
 GN SHK2 OR SPACIF5.09C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98325061; PubMed-9660817;
RA Yang P., Kanstra S., Pimental R.A., Gilbreth M., Marcus S.;
RT "Cloning and characterization of shk2, a gene encoding a novel p21-
RT activated protein kinase from fission yeast.";
RL J. Biol. Chem. 273:18481-18489(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Gentles S., Odell C., Churcher C.M., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FORMS AN ACTIVATED COMPLEX WITH GTP-BOUND RAS-LIKE
CC CD42. PARTICIPATES IN RAS-DEPENDENT MORPHOLOGICAL CONTROL AND
CC MARTING RESPONSE PATHWAYS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 GBD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U45981; AAA87575.1; -
DR DB: 268136; CAA92237.1; -
DR HSP: P00518; LPHK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000095; PAK_box_P21_Rho_binding.
DR InterPro: IPR001849; PH.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00786; PBD; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00069; pkinase; 1.
DR SMART: SM00285; PBD; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS50108; GBD; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 23 125
FT DOMAIN 129 186
FT DOMAIN 309 566
FT NP_BIND 315 323
FT BINDING 343 343
FT ACT_SITE 434 434
FT ACT_SITE 434 434
SQ SEQUENCE 589 AA: 66765 MW: CABE2190925EC31 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 589;
Best Local Similarity 42.9%; Pred. No. 62;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
OY 1 LVRRFVHRRPHVES 14
DB 251 IIRPTEKSHVES 264

RESULT 12
GLMS_BACHD STANDARD; PRT; 599 AA.
AC Q9KG45;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)

```

```

DE Glucosamine-6-phosphate aminotransferase (isomerizing)
DE (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-
DE phosphate amidotransferase) (GPA) (L-glutamine-D-fructose-6-phosphate
DE amidotransferase) (Glucosamine-6-phosphate synthase).
GN GLMS OR BH0268.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE-20512582; PubMed-11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.",
RT Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: Catalyzes the first step in hexosamine metabolism,
CC converting fructose-6P into glucosamine-6P using glutamine as a
CC nitrogen source (By similarity).
CC -1- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate -> L-
CC glutamate + D-glucosamine 6-phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2
CC GATASE DOMAIN.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.
CC GLMS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP01507; BAB03987.1; -
DR HSP: P17169; IMOG.
DR InterPro: IPR000583; GATase_2.
DR InterPro: IPR001347; SIS.
DR Pfam: PF00310; GATase_2; 1.
DR Pfam: PF01380; SIS; 2.
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
KW Complete proteome.
FT INIT_MET 0 0
FT DOMAIN 1 239
FT ACT_SITE 594 594
FT ACT_SITE 594 594
SQ SEQUENCE 599 AA: 65746 MW: DE9C2667EA70C285 CRC64;

Query Match 46.8%; Score 37; DB 1; Length 599;
Best Local Similarity 57.1%; Pred. No. 63;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
OY 1 LVRRFVHRRPHVES 14
DB 129 LVRRFVHRRPHVES 142

RESULT 13
AMYA_PYRFU STANDARD; PRT; 648 AA.
AC P49067;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PF0272.
OS Pyrococcus furiosus.

```

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 ON NCBI\_TaxID=2261;  
 RX [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=94043280; PubMed=8226990.  
 RA Laderman K.A., Asada K., Uemori T., Mukai H., Taguchi Y., Kato I.,  
 RA Anfinsen C.B.;  
 RT "Alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus  
 RT furiosus. Cloning and sequencing of the gene and expression in  
 RT Escherichia coli";  
 RL J. Biol. Chem. 268:24402-24407(1993).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94043279; PubMed=8226989;  
 RA Laderman K.A., Davis B.R., Kruttsch H.C., Lewis M.S., Griko Y.V.,  
 RA Privetov P.L., Anfinsen C.B.;  
 RT "The purification and characterization of an extremely thermostable  
 RT alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus  
 RT furiosus";  
 RL J. Biol. Chem. 268:24394-24401(1993).  
 CC -1- FUNCTION: DISPLAYS A BROAD RANGE OF SUBSTRATE SPECIFICITY. WITH  
 CC THE CAPACITY TO HYDROLYSE CARBOHYDRATES AS SIMPLE AS MALTRIOSE.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic  
 CC linkages in oligosaccharides and polysaccharides.  
 CC -1- PATHWAY: POLYSACCHARIDE DEGRADATION.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- MISCELLANEOUS: THE ISOELECTRIC POINT IS 4.3. THE ENZYME DISPLAYS  
 CC OPTIMAL ACTIVITY, WITH SUBSTANTIAL THERMAL STABILITY, AT 100  
 CC DEGREES CELSIUS, WITH THE ONSET OF ACTIVITY AT APPROXIMATELY 40  
 CC DEGREES CELSIUS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L22346; AAA72035.1; -  
 DR InterPro: IPR004300; Glyco\_hydro\_57.  
 DR Pfam: PF03065; Glyco\_hydro\_57; 1.  
 KW Hydrolyase; Glycosidase; Carbohydrate metabolism.  
 FT INIT\_MET 0  
 FT SEQUENCE 648 AA; 76178 MW; BF7A495F084E0FB1 CRC64;  
 SQ  
 Query Match 46.8%; Score 37; DB 1; Length 648;  
 Best Local Similarity 53.3%; Pred. No. 68;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 LVRRFVHRPPVESQ 15  
 DB 436 LARMEHYHGVESQ 450  
 RESULT 14  
 BROW\_DROME STANDARD; PRT: 675 AA.  
 AC P12428; Q24264;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Brown protein.  
 GN BW.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=89218981; PubMed=3149712;  
 RA Dreesen T.D., Johnson D.H., Henikoff S.;  
 RT "The brown protein of Drosophila melanogaster is similar to the white  
 RT protein and to components of active transport complexes";  
 RL Mol. Cell. Biol. 8:5206-5215(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95044127; PubMed=7956072;  
 RA Martin-Morris L.E., Loughney K., Kershinsnik E.O., Poorlinga G.,  
 RA Henikoff S.;  
 RT "Characterization of sequences responsible for trans-inactivation of  
 RT the Drosophila brown gene";  
 RL Cold Spring Harb. Symp. Quant. Biol. 58:577-584(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=10281;  
 RA Nitasaka E., Green M.W., Yamazaki T.;  
 RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY  
 CC FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS  
 CC RESPONSIBLE FOR EYE COLOR. BROWN AND WHITE DIMERIZE FOR THE  
 CC TRANSPORT OF GUANINE.  
 CC -1- SUBUNIT: HETERODIMER OF BROWN WITH WHITE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M20630; AAA28397.1; -  
 DR EMBL: L23543; AAC37214.1; -  
 DR EMBL: L05635; AAA28398.1; -  
 DR PIR: A31399; FYFFB.  
 DR FLYBase: FBgn0000241; bw.  
 DR InterPro: IPR003439; ABC\_transportr.  
 DR InterPro: IPR001687; ATP\_GTP\_A.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP\_BIND 66 73  
 FT TRANSMEM 423 441  
 FT TRANSMEM 454 474  
 FT TRANSMEM 504 522  
 FT TRANSMEM 531 552  
 FT TRANSMEM 565 583  
 FT TRANSMEM 645 665  
 FT TRANSMEM 28 28  
 FT VARIANT 274 274  
 FT VARIANT 331 331  
 FT VARIANT 407 407  
 FT VARIANT 638 638  
 FT CONFLICT 44 44  
 FT SEQUENCE 675 AA; 75943 MW; 81DEBDE856FAF174 CRC64;  
 SQ  
 Query Match 46.8%; Score 37; DB 1; Length 675;  
 Best Local Similarity 75.0%; Pred. No. 71;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 LVRRFVHR 8  
 DB 664 LVRRYIHR 671  
 RESULT 15  
 E2K3\_RAT STANDARD; PRT: 1108 AA.  
 ID E2K3\_RAT  
 AC Q92121;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Eukaryotic translation initiation factor 2-alpha kinase 3 precursor  
 DE (EC 2.7.1.-) (PKR-like endoplasmic reticulum kinase) (Pancreatic  
 DE eIF2-alpha kinase).  
 GN EIF2AK3 OR PERK OR PERK.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N. A., AND CHARACTERIZATION.  
 RC TISSUE-Pancreatic islets; PubMed=9819435;  
 RX MEDLINE=99038253; PubMed=9819435;  
 RA Shi Y., Valtem K.M., Sood R., An J., Liang J., Stramm L.E., Wek R.C.;  
 RT "Identification and characterization of pancreatic eukaryotic  
 RT initiation factor 2 alpha-subunit kinase, PERK, involved in  
 RT translational control".  
 RL Mol. Cell. Biol. 18:7499-7509(1998).  
 RN [2]  
 RP MUTAGENESIS OF LYS-614.  
 RX MEDLINE=99150360; PubMed=10026192;  
 RA Shi Y., An J., Liang J., Hayes S.E., Sandusky G.E., Stramm L.E.,  
 RA Yang N.N.;  
 RT "Characterization of a mutant pancreatic eIF-2alpha kinase, PERK, and  
 RT co-localization with somatostatin in islet delta cells.";  
 RL J. Biol. Chem. 274:5723-5730(1999).  
 RN [3]  
 RP SUBUNIT.  
 RX MEDLINE=20313073; PubMed=10854322;  
 RA Bertolotti A., Zhang Y., Hendershot L.M., Harding H.P., Ron D.;  
 RT "Dynamic interaction of BiP and ER stress transducers in the  
 RT unfolded-protein response.";  
 RL Nat. Cell Biol. 2:326-332(2000).  
 CC -1- FUNCTION: PHOSPHORYLATES THE ALPHA SUBUNIT OF EUKARYOTIC  
 CC TRANSLATION-INITIATION FACTOR 2 (EIF2), LEADING TO ITS  
 CC INACTIVATION AND THUS TO A RAPID REDUCTION OF TRANSLATIONAL  
 CC INITIATION AND REPRESSION OF GLOBAL PROTEIN SYNTHESIS. SERVES AS A  
 CC CRITICAL EFFECTOR OF UNFOLDED PROTEIN RESPONSE (UPR)-INDUCED G1  
 CC GROWTH ARREST DUE TO THE LOSS OF CYCLIN D1 (BY SIMILARITY).  
 CC -1- ENZYME REGULATION: PERTURBATION IN PROTEIN FOLDING IN THE  
 CC ENDOPLASMIC RETICULUM (ER) PROMOTES REVERSIBLE DISSOCIATION FROM  
 CC HSP45/BIP AND OLIGOMERIZATION, RESULTING IN  
 CC TRANSAUTOPHOSPHORYLATION AND KINASE ACTIVITY INDUCTION.  
 CC -1- SUBUNIT: FORMS DIMERS WITH HSP45/BIP IN RESTING CELLS.  
 CC OLIGOMERIZES IN ER-STRESSED CELLS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic  
 CC reticulum.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
 CC -1- INDUCTION: BY ER STRESS.  
 CC -1- DOMAIN: THE LUMENAL DOMAIN SENSES PERTURBATIONS IN PROTEIN FOLDING  
 CC IN THE ER, PROBABLY THROUGH REVERSIBLE INTERACTION WITH HSP45/BIP.  
 CC -1- PTM: AUTOPHOSPHORYLATED.  
 CC -1- PTM: N-GLYCOSYLATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC GCN2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF096835; AAC83801.1; -;  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00069; pkinase; 2.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE: PS0011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.

KW Serine/threonine-protein kinase; Transferase; ATP-binding;  
 KW Translation regulation; Unfolded protein response;  
 KW Endoplasmic reticulum; Phosphorylation; Glycoprotein; Signal;  
 KW Transmembrane.  
 FT SIGNAL 1 27  
 FT CHAIN 28 1108  
 FT  
 FT DOMAIN 28 506  
 FT TRANSMEM 507 527  
 FT DOMAIN 528 1108  
 FT DOMAIN 585 1069  
 FT NP\_BIND 591 599  
 FT BINDING 614 614  
 FT ACT\_SITE 929 929  
 FT DOMAIN 47 50  
 FT DOMAIN 223 228  
 FT CARBOHYD 253 253  
 FT MUTAGEN 614 614  
 SQ SEQUENCE 1108 AA: 124769 MW: 83716 BIPD26ED32B CRC64;

Query Match 46.8%; Score 37; DB 1; Length 1108;  
 Best Local Similarity 54.5%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 LVRRFVHRPH 11  
 :||| |:  
 Db 531 IYRRLFPQPH 541

Search completed: August 15, 2002, 11:47:42  
 Job time: 518 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:47:17 ; Search time 58.4 Seconds  
(without alignments)  
44.434 Million cell updates/sec

Title: US-09-613-092a-7  
Perfect score: 79  
Sequence: 1 LVRRFVRRPHEVQ 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	52	65.8	243	5	09VRG7	09VRG7	drosophila
2	52	65.8	244	4	095476	095476	homo sapien
3	52	65.8	244	4	096G09	096G09	homo sapien
4	46	58.2	324	16	P94387	P94387	mus musculus
5	45	57.0	577	16	09Z9B5	09Z9B5	chlamydia p
6	45	57.0	798	5	09U4H1	09U4H1	drosophila
7	45	57.0	798	5	09V866	09V866	drosophila
8	44	55.7	546	11	061491	061491	mus musculus
9	43	54.4	72	9	064123	064123	bacterioph
10	43	54.4	72	16	031924	031924	streptomyces
11	42	53.2	325	2	09F2P9	09F2P9	mus musculus
12	41	51.9	168	11	09D9D6	09D9D6	caenorhabditis
13	41	51.9	254	5	095XU1	095XU1	caenorhabditis
14	41	51.9	466	3	000086	000086	talatomyces
15	41	51.9	481	5	017558	017558	caenorhabditis
16	40	50.6	138	17	09YBN3	09YBN3	aeropyrum p

17	40	50.6	439	2	09Z4R6	09Z4R6	eikenella c
18	40	50.6	542	5	0960K4	0960K4	drosophila
19	40	50.6	1087	5	09B104	09B104	caenorhabditis
20	40	50.6	1220	12	039272	039272	equine herpes
21	40	50.6	1233	11	P97820	P97820	mus musculus
22	40	50.6	1255	5	095VU9	095VU9	drosophila
23	40	50.6	1317	5	09VLK6	09VLK6	drosophila
24	39	49.4	67	3	004334	004334	pichia angu
25	39	49.4	80	16	09YXH1	09YXH1	neisseria m
26	39	49.4	148	10	004427	004427	chlamydomon
27	39	49.4	181	4	096N68	096N68	homo sapien
28	39	49.4	183	16	09ZKK8	09ZKK8	rhizobium m
29	39	49.4	232	12	099GPO	099GPO	culex nigri
30	39	49.4	246	5	020432	020432	caenorhabditis
31	39	49.4	257	16	066673	066673	aquifex aco
32	39	49.4	261	3	09P390	09P390	neurospora
33	39	49.4	409	5	09VND6	09VND6	drosophila
34	39	49.4	410	5	0906M0	0906M0	drosophila
35	39	49.4	476	3	059718	059718	schizosacch
36	39	49.4	540	4	092783	092783	homo sapien
37	39	49.4	548	11	P70297	P70297	mus musculus
38	39	49.4	564	10	09FR00	09FR00	oryza sativ
39	39	49.4	954	16	084594	084594	chlamydia t
40	39	49.4	1207	16	09A797	09A797	caulobacter
41	39	49.4	1792	5	097417	097417	drosophila
42	39	49.4	1792	5	09V4M6	09V4M6	drosophila
43	39	49.4	1927	5	09GY95	09GY95	leishmania
44	38	48.1	141	5	09VD20	09VD20	drosophila
45	38	48.1	217	5	095TP9	095TP9	drosophila

## ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	09VRG7	PRELIMINARY; PRT; 243 AA.
AC	09VRG7	
DT	01-MAY-2000 (TREMURel. 13, Created)	
DT	01-MAY-2000 (TREMURel. 13, Last sequence update)	
DE	01-DEC-2001 (TREMURel. 19, Last annotation update)	
DE	CG1696 PROTEIN.	
GN	CG1696.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams W.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Adair J.F., Agbayani A., An H.-J., Andrews-Plankoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Besono K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Paolo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,	
RA	Hoskins R., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,	
RA	Jatani M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mervulov G., Milshina N.V., Mobarry C., Morris J., Mostreli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Messarum D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003568; AAF50833.1; -  
 DR FLYBASE: FBgn0031173; CG1696.  
 DR InterPro: IPR004274; NIF:  
 DR Pfam: PF03031; NIF:  
 SQ SEQUENCE 243 AA; 28486 MW; 1E278DD1D8DF60C6 CRC64;

Query Match 65.8%; Score 52; DB 5; Length 243;  
 Best Local Similarity 75.0%; Pred. No. 0.33;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VRFVHRRPHE 13  
 |||||:  
 DB 103 VRFVHRRPHE 114

RESULT 2  
 ID 095476 PRELIMINARY; PRT; 244 AA.  
 AC 095476:  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL. 28.3 KDA PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Keen J., Inghearn C.;  
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ011916; CA09865.1; -  
 DR InterPro: IPR004274; NIF:  
 DR Pfam: PF03031; NIF: 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 244 AA; 28347 MW; E23843B815DEFDF0 CRC64;

Query Match 65.8%; Score 52; DB 4; Length 244;  
 Best Local Similarity 75.0%; Pred. No. 0.33;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VRFVHRRPHE 13  
 |||||:  
 DB 104 VRFVHRRPHE 115

RESULT 3  
 ID 096609 PRELIMINARY; PRT; 244 AA.  
 AC 096609:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE UNKNOWN (PROTEIN FOR MGC:16648).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;  
 RA Strausberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC009295; AAH09295.1; -  
 SQ SEQUENCE 244 AA; 28377 MW; 062952A90F74575A CRC64;

Query Match 65.8%; Score 52; DB 4; Length 244;  
 Best Local Similarity 75.0%; Pred. No. 0.33;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VRFVHRRPHE 13  
 |||||:  
 DB 104 VRFVHRRPHE 115

RESULT 4  
 ID P94387 PRELIMINARY; PRT; 324 AA.  
 AC P94387:  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HOMOLOGUE OF REGULATORY PROTEIN OXFR OF E. COLI.  
 GN YCGK.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168.  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummins N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Eutlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Iraya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-blanchard M., Klein C.,  
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Moone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takekura K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni K.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler H., Weitzengger T.,  
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*  
 RT *subtilis*,"  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-168:  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
 DR EMBL: D50453; BAA08951.1; -;  
 DR EMBL: Z99105; CAB2111.1; -;  
 DR InterPro: IPR000847; HTH\_LySR.  
 DR Pfam: PF00126; HTH\_1; 1.  
 DR PRINTS: PR00039; HTHLYSR.  
 DR PROSITE: PS00044; HTH\_LYSR\_FAMILY; UNKNOWN\_1.  
 KM Complete proteome: DNA-binding; Transcription regulation.  
 SO SEQUENCE 324 AA; 36406 MW; C0C8FD25C3E08D97 CRC64;

Query Match 58.2%; Score 46; DB 16; Length 324;  
 Best Local Similarity 60.0%; Pred. No. 4.6;  
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LVRFVHRPHVESQ 15  
 Db 109 LKQFVARYPHVEVO.123

RESULT 5  
 0929B5 PRELIMINARY; PRT; 577 AA.  
 AC 0929B5;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOHETICAL 65.1 KDA PROTEIN.  
 GN CPN0066 OR CPJ0066 OR CP0708.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE=99206606; Pubmed-10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis,"  
 RL Nf. Genet. 21:385-389(1999).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-AR39;  
 RX MEDLINE=20150259; Pubmed-10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gayn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39,"  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-J138;  
 RX MEDLINE=20330349; Pubmed-10871362;  
 RA Shira T., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA,"  
 RL Nucleic Acids Res. 28:2311-2314(2000).

DR EMBL: AE001592; AAD18219.1; -;  
 DR EMBL: AE002229; AAF38516.1; -;  
 DR EMBL: AP002545; BAA98277.1; -;  
 DR TIGR: CP0708; -;  
 KM Hypothetical protein; Complete proteome.  
 SO SEQUENCE 577 AA; 65110 MW; FC436FA530116A8F CRC64;

Query Match 57.0%; Score 45; DB 16; Length 577;  
 Best Local Similarity 80.0%; Pred. No. 12;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VRFVHRPH 11  
 Db 567 VRRRLHRRPH 576

RESULT 6  
 0904H1 PRELIMINARY; PRT; 798 AA.  
 AC 0904H1;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE BCDNA:GH05095.  
 GN BCDNA:GH05095 OR CG6424.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rudin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,  
 RA Agbayan A., Arcalata T.T., Baxter E., Blazej R.G., Butenhof C.,  
 RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,  
 RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,  
 RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,  
 RA Park S., Sequelira A., Sethi H., Snir E., Svitskas R.R., Weinburg T.,  
 RA Celinker S.E.;  
 RT "Full Length Drosophila melanogaster cDNA sequence,"  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF181632; AAD5418.1; -;  
 DR FlyBase: FBgn0028494; BCDNA:GH05095.  
 SO SEQUENCE 798 AA; 88467 MW; OD1B9DB99243752A CRC64;

Query Match 57.0%; Score 45; DB 5; Length 798;  
 Best Local Similarity 58.3%; Pred. No. 17;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LVRFVHRPHV 12  
 Db 374 LSRRAHSRPHI 385

RESULT 7  
 09V866 PRELIMINARY; PRT; 798 AA.  
 AC 09V866;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE BCDNA:GH05095 PROTEIN.  
 GN BCDNA:GH05095 OR CG6424.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;



```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchett S.,
RA Borriis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabre C., Ferrari E., Foidger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holstrel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klauer-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Presseran E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpsita P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartot A.,
RA Vlati A., Wandt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: 299114; CAB13949.1; -.
KW Complete proteome.
SQ SEQUENCE 72 AA; 8382 MW; 11178AF395985ED8 CRC64;

Query Match 54.4%; Score 43; DB 16; Length 72;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 FVHRPH 11
Db 60 FVHRPH 66

RESULT 11
Q9F2P9 PRELIMINARY; PRT; 325 AA.
ID Q9F2P9;
AC Q9F2P9;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN SCE41.09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denaplate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL442120; CAC09541.1; -.
KW Hypothetical protein.
SQ SEQUENCE 325 AA; 34793 MW; 993E789D3B2ADD2C CRC64;

Query Match 53.2%; Score 42; DB 2; Length 325;
Best Local Similarity 58.3%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LVRFVHRPHV 12
Db 211 LVDRKLVHRPHV 222

RESULT 12
Q9D9D6 PRELIMINARY; PRT; 168 AA.
ID Q9D9D6;
AC Q9D9D6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 1700095D18RIK PROTEIN.
GN 1700095D18RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Baren G.,
RA Blake J., Botfield D., Boujanga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wells C., Whitaker C., Wilming L.,
RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK007076; BAB24851.1; -.
DR MGD; MGI:1920805; 1700095D18RIK.
DR InterPro; IPR000504; RRM.
DR PROSITE; PS00030; RRM_1; UNKNOWN_1.
SQ SEQUENCE 168 AA; 17964 MW; 144EA0BA9238359 CRC64;

Query Match 51.9%; Score 41; DB 11; Length 168;
Best Local Similarity 50.0%; Pred. No. 17;

```

```

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Oy 3 RRFVHRPPHVES 14
    1:1:1 11:1
Db 65 RQFHSLSLPHNQS 76

RESULT 13
O95XU1 PRELIMINARY; PRT; 254 AA.
AC 095XU1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 27.6 KDA PROTEIN.
GN Y67D8A.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None.
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Edwards J., Lamar B., Mlux P., Du H., Kemp K., Woldmann P.,
Walker C.;
RT "The sequence of C. elegans cosmid Y67D8A."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC024848; AK68545.1; -.
KW Hypothetical protein
SQ SEQUENCE 254 AA; 27600 MW; 32895C8BD425785 CRC64;

Query Match 51.9%; Score 41; DB 5; Length 254;
Best Local Similarity 70.0%; Pred. No. 26;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 1 LVRRFVHRPP 10
    1:111 11:1
Db 71 LVRRFHRNP 80

RESULT 14
O00096 PRELIMINARY; PRT; 466 AA.
AC 000096;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PHYBASE (EC 3.1.3.8).
OS Talaromyces thermophilus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Talaromyces.
OX NCBI_TaxID=28565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98007872; PubMed=9349716;
RA Pasamontes L., Halker M., Henriquez-Huecas M., Mitchell D.B.,
van Loon A.P.;
RT "Cloning of the phytases from Emericella nidulans and the thermophilic

```

```

RT fungus Talaromyces thermophilus."
RL Biochim. Biophys. Acta 1353:217-223(1997).
DR EMBL: U59802; AAB96873.1; -.
DR HSSP: P34752; 1HP.
DR InterPro: IPR000560; His_acid_phosphatase.
DR Pfam: PF00328; acid_phosphat_1.
DR PROSITE: PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE: PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 466 AA; 51450 MW; FC4575B521A5C929 CRC64;

Query Match 51.9%; Score 41; DB 3; Length 466;
Best Local Similarity 57.1%; Pred. No. 47;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Oy 1 LVRRFVHRPPHVES 14
    1:1:1 11:1
Db 12 LVALLVSRNPHVS 25

RESULT 15
O17558 PRELIMINARY; PRT; 481 AA.
ID 017558;
AC 017558;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COLF.1 PROTEIN.
GN COLF.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,
Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
Smaison N., Smith A., Sönnhammer E., Staden R., Sulston J.,
Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Johnson D.;
RT "The sequence of C. elegans cosmid COLF.1."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U58761; AAB00717.1; -.
SQ SEQUENCE 481 AA; 54463 MW; 00FE20C409BCC9A CRC64;

Query Match 51.9%; Score 41; DB 5; Length 481;
Best Local Similarity 42.9%; Pred. No. 49;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Oy 2 VRRFVHRPPHVESQ 15
    111:1 11:1
Db 416 VRRYLRKRPPTKE 429

```

• Fri Aug 16 10:36:06 2002

us-09-613-092a-7.rspt

Page 7

Search completed: August 15, 2002, 11:47:18  
Job time: 529 sec

---

**THIS PAGE BLANK (USPTO)**



Fri Aug 16 10:36:40 2002

devi-613-comb.rag

Page 1

Seg. 1D. 5 (1-15) + 6 (1-15) + 7 (1-15)  
residue residue residue

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:50:25 ; Search time 51.96 Seconds  
(Without alignments)  
96.196 Million cell updates/sec

Title: DEVI-613-COMB  
Perfect score: 257  
Sequence: 1 TVSRVPTMAAFHGYSYQH.....GFWLLVRFRVRRPHVESQ 45

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	35.4	15	20	AAV30351
2	87	33.9	15	20	AAV30352
3	79	30.7	15	20	AAV30353
4	74	28.8	15	20	AAV30354
5	63	24.5	257	21	AAV68864
6	63	24.5	281	21	AAW93592
7	63	24.5	281	21	AAV93213
8	58.5	22.8	398	20	AAW78169
9	58.5	22.8	404	22	AAU14277
10	56.5	22.0	459	21	AAV5223
11	56.5	22.0	459	21	AAV5225

12	56	21.8	214	21	AAV93216
13	56	21.8	229	21	AAV68865
14	56	21.8	242	21	AAV93214
15	56	21.8	731	22	AAV91100
16	56	21.8	731	22	AAV9423
17	55	21.4	43	21	AAV19038
18	55	21.4	125	21	AAV19037
19	55	21.4	285	19	AAW60736
20	54.5	21.2	422	22	AAV33276
21	54.5	21.2	422	22	AAV6548
22	54.5	21.2	500	22	AAV92760
23	54.5	21.2	951	20	AAV34536
24	54.5	21.2	953	20	AAV34403
25	54	21.0	117	22	AAV6510
26	54	21.0	631	22	ABV63273
27	54	21.0	631	22	AAU38950
28	53.5	20.8	1182	22	ABV25897
29	53.5	20.8	1216	22	AAE08076
30	53.5	20.8	1247	22	AAE08077
31	53.5	20.8	1332	22	AAE08078
32	53.5	20.8	1363	22	AAE08075
33	53	20.6	138	22	AAV75185
34	53	20.6	197	22	ABV08246
35	53	20.6	197	22	AAV47135
36	53	20.6	215	18	AAV17522
37	52	20.2	66	19	AAV7616
38	52	20.2	156	22	AAV75292
39	52	20.2	156	22	AAV75293
40	52	20.2	224	22	AAU34219
41	52	20.2	230	22	AAU36882
42	52	20.2	243	22	ABV58399
43	52	20.2	273	22	ABV18966
44	52	20.2	280	22	ABV10923
45	52	20.2	282	22	ABV10922

#### ALIGNMENTS

RESULT 1  
AAV30351  
ID AAV30351 standard; Peptide: 15 AA.  
XX  
AC AAV30351;  
XX  
DF 09-NOV-1999 (first entry)  
XX  
DE Epitope derived from pneumococcal surface adhesion A protein.  
XX  
DE Pneumococcal surface adhesion A protein; Peaa; monoclonal antibody;  
KW vaccine; Streptococcus pneumoniae infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN W09945121-A1.  
XX  
PD 10-SEP-1999.  
XX  
PF 26-FEB-1999; 99WO-US04326.  
XX  
PR 02-MAR-1998; 98US-0076565.  
XX  
(USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;  
PI Zeller JL;  
XX  
XX WPI; 1999-540849/45.  
XX  
XX New peptides corresponding to Streptococcus pneumoniae Peaa, used  
XX for treating or preventing Streptococcus pneumoniae infection in a  
XX subject

Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
C glutamic prote  
Corynebacterium gl  
Zea mays protein f  
Zea mays protein f  
Amino acid sequenc  
Pinus radiata tran  
Corynebacterium gl  
C glutamic prote  
Porphyromonas ging  
Porphyromonas ging  
Human reproductive  
Drosophila melanog  
Drosophila G-prote  
Novel human diagno  
Human transporter-  
Human transporter-  
Human transporter-  
Human transporter-  
Human colon cancer  
Novel human diagno  
CD1E-16, Incyte I  
Human beta-A3-crys  
Staphylococcus aur  
Gene 6 human secre  
Human secreted pro  
Staphylococcus aur  
Staphylococcus aur  
Drosophila melanog  
Novel human diagno  
Novel human diagno  
Novel human diagno

```

PS Claim 6; Page 43; 58pp; English.
XX
CC AAY30351-54 represent immunogenic peptides which are derived from
CC a pneumococcal surface adhesion A protein (PsaA). The specification
CC describes monoclonal antibodies which bind epitopes of the PsaA protein
CC (e.g present sequence). The peptides can be used in vaccines to prevent
CC Streptococcus pneumoniae infections. The antibodies of the invention
CC can also be used to detect S. pneumoniae in a sample or individual.
XX
SQ Sequence 15 AA;

Query Match 35.4%; Score 91; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TVSRVPMWTAMAFHGY 15
   |||||
DB 1 tvsrvpwtawafngy 15

RESULT 2
AAY30352
ID AAY30352 standard; Peptide; 15 AA.
XX
AC AAY30352;
XX
DT 09-NOV-1999 (first entry)
XX
DE Epitope derived from pneumococcal surface adhesion A protein.
XX
KM Pneumococcal surface adhesion A protein; PsaA; monoclonal antibody;
KM vaccine; Streptococcus pneumoniae infection.
XX
OS Streptococcus pneumoniae.
XX
PN WO9945121-A1.
XX
PD 10-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04326.
XX
PR 02-MAR-1998; 98US-0076565.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;
PI Zeiler JL;
XX
DR WPI; 1999-540849/45.
XX
PT New peptides corresponding to Streptococcus pneumoniae PsaA, used
PT for treating or preventing Streptococcus pneumoniae infection in a
PT subject
XX
PS Claim 6; Page 43; 58pp; English.
XX
CC AAY30351-54 represent immunogenic peptides which are derived from
CC a pneumococcal surface adhesion A protein (PsaA). The specification
CC describes monoclonal antibodies which bind epitopes of the PsaA protein
CC (e.g present sequence). The peptides can be used in vaccines to prevent
CC Streptococcus pneumoniae infections. The antibodies of the invention
CC can also be used to detect S. pneumoniae in a sample or individual.
XX
SQ Sequence 15 AA;

Query Match 33.9%; Score 87; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 RSYQHDLRAYGFWR 30
   |||||

```

```

DB 1 rsyqhdlraygfwr 15

RESULT 3
AAY30353
ID AAY30353 standard; Peptide; 15 AA.
XX
AC AAY30353;
XX
DT 09-NOV-1999 (first entry)
XX
DE Epitope derived from pneumococcal surface adhesion A protein.
XX
KM Pneumococcal surface adhesion A protein; PsaA; monoclonal antibody;
KM vaccine; Streptococcus pneumoniae infection.
XX
OS Streptococcus pneumoniae.
XX
PN WO9945121-A1.
XX
PD 10-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04326.
XX
PR 02-MAR-1998; 98US-0076565.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;
PI Zeiler JL;
XX
DR WPI; 1999-540849/45.
XX
PT New peptides corresponding to Streptococcus pneumoniae PsaA, used
PT for treating or preventing Streptococcus pneumoniae infection in a
PT subject
XX
PS Claim 6; Page 43; 58pp; English.
XX
CC AAY30351-54 represent immunogenic peptides which are derived from
CC a pneumococcal surface adhesion A protein (PsaA). The specification
CC describes monoclonal antibodies which bind epitopes of the PsaA protein
CC (e.g present sequence). The peptides can be used in vaccines to prevent
CC Streptococcus pneumoniae infections. The antibodies of the invention
CC can also be used to detect S. pneumoniae in a sample or individual.
XX
SQ Sequence 15 AA;

Query Match 30.7%; Score 79; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 LVRRFVHRHRYVESQ 45
   |||||
DB 1 lvrrfvhrtrphvesq 15

RESULT 4
AAY30354
ID AAY30354 standard; Peptide; 15 AA.
XX
AC AAY30354;
XX
DT 09-NOV-1999 (first entry)
XX
DE Epitope derived from pneumococcal surface adhesion A protein.
XX
KM Pneumococcal surface adhesion A protein; PsaA; monoclonal antibody;
KM vaccine; Streptococcus pneumoniae infection.
XX
OS Streptococcus pneumoniae.
XX

```

PN W09945121-A1.  
 XX  
 PD 10-SEP-1999.  
 XX  
 PF 26-FEB-1999; 99WO-US04326.  
 XX  
 PR 02-MAR-1998; 98US-0076565.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Ades EW, Carlone GW, Sampson JS, Tharpe JA, Westerink MAJ;  
 PI Zeller JI;  
 DR WPI: 1999-540849/45.  
 XX  
 PT New peptides corresponding to Streptococcus pneumoniae Psaa, used  
 PT for treating or preventing Streptococcus pneumoniae infection in a  
 PT subject  
 XX  
 PS Claim 6; Page 43; 58pp; English.  
 XX  
 CC AAY30351-54 represent immunogenic peptides which are derived from  
 CC a pneumococcal surface adhesion A protein (psaa). The specification  
 CC describes monoclonal antibodies which bind epitopes of the psaa protein  
 CC (e.g present sequence). The peptides can be used in vaccines to prevent  
 CC Streptococcus pneumoniae infections. The antibodies of the invention  
 CC can also be used to detect S. pneumoniae in a sample or individual.  
 XX  
 SQ Sequence 15 AA:

Query Match 28.8%; Score 74; DB 20; Length 15;  
 Best Local Similarity 93.3%; Pred. No. 0.0013;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 31 LVRRFVRRPVEQ 45  
 |||||  
 Db 1 LVRRFVRRPVEQ 15

RESULT 5  
 AAY68864  
 ID AAY68864 standard; Protein; 257 AA.  
 XX  
 AC AAY68864;  
 XX  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE Amino acid sequence of a murine caspase-like polypeptide.  
 XX  
 KW Mouse; caspase-like polypeptide; human; caspase; apoptosis;  
 KW skin disease; keratinisation; wound healing.  
 XX  
 OS Mus musculus.  
 XX  
 PN W020004169-A1.  
 PD 27-JAN-2000.  
 XX  
 PF 12-JUL-1999; 99WO-EP04939.  
 XX  
 PR 17-JUL-1998; 98EP-0202422.  
 XX  
 PA (VLAAR-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
 XX  
 PI Van De Craen M, Declercq W, Vandenaabeele P, Fliers W;  
 XX  
 DR WPI: 2000-182433/16.  
 DR N-PSDB; AAZ60683.  
 XX  
 PT New murine and human caspase homologues useful for treating skin  
 PT related disorders  
 XX

PS Claim 1; Page 51-52; 68pp; English.  
 XX  
 CC The present sequence represents a murine caspase-like polypeptide. The  
 CC specification also describes a human caspase-like polypeptide.  
 CC Caspases are cysteinyl aspartate-specific proteinases which play a  
 CC central role in apoptosis. The polypeptides of the invention are related  
 CC to human and murine caspase-2 and human caspase-9, and possess all of  
 CC the typical amino acids involved in catalysis, including the QACRG box,  
 CC and contain no or only a very short prodomain. mRNA expression of the  
 CC homologues of the invention is predominant in the skin. The caspase-like  
 CC polypeptides are useful for treating human or animal diseases, such  
 CC as skin diseases. They are also useful for screening for compounds that  
 CC modulate its activity, i.e. agonists, antagonists, and inhibitors. The  
 CC caspase-like polypeptides and polynucleotides are useful for modulating  
 CC keratinisation, for diagnosing and treating inappropriate wound  
 CC healing.  
 XX  
 SQ Sequence 257 AA:

Query Match 24.5%; Score 63; DB 21; Length 257;  
 Best Local Similarity 42.3%; Pred. No. 1;  
 Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

OY 14 GYRSYQHDLRAYGFWRLLVRRFVRR 39  
 ||||| : || : | 1:| :  
 Db 189 gylsyrhdxsgfqlgtldvfhkx 214

RESULT 6  
 AAW93592  
 ID AAW93592 standard; Protein; 281 AA.  
 XX  
 AC AAW93592;  
 XX  
 DT 21-JUN-1999 (first entry)  
 XX  
 DE Mouse caspase-14 protein.  
 XX  
 KW Caspase-14; murine; protease; treatment; apoptotic-related disease;  
 KW autoimmune disease; cancer; acquired immunodeficiency syndrome; AIDS;  
 KW neurodegenerative disease; ischaemic injury; anti-idiotypic antibody;  
 KW caspase-14 processing activity; epitope; competitor; modulator.  
 XX  
 OS Mus sp.  
 XX  
 PN W09910504-A2.  
 PD 04-MAR-1999.  
 XX  
 PF 26-AUG-1998; 98WO-US17715.  
 XX  
 PR 26-AUG-1997; 97US-0056986.  
 XX  
 PA (IDUN-) IDUN PHARM INC.  
 XX  
 PI Alnemrl ES, Fernandes-Alnemrl T;  
 XX  
 DR WPI: 1999-204670/17.  
 DR N-PSDB; AAX23515.  
 XX  
 PT Newly isolated polynucleotide encoding a caspase-14 polypeptide  
 PT useful for identifying (ant)agonists that are useful in the  
 PT diagnosis and treatment of apoptosis-related diseases  
 XX  
 PS Claim 4; Fig 1; 59pp; English.  
 XX  
 CC This invention describes a novel murine caspase-14 which has protease  
 CC activity. The caspase-14 polypeptide is useful for identifying  
 CC (ant)agonists of the polypeptide, where enzyme activity is measured  
 CC with a fluorescent substrate (especially DEVD-AMC or IVAD-AMC. Activated  
 CC caspase-14 is useful for identifying inhibitors or enhancers of  
 CC caspase-14 activity. The compounds identified by both methods

CC form pharmaceutical compositions for treating apoptotic-related diseases, including autoimmune disease, cancer, acquired immunodeficiency syndrome (AIDS), neurodegenerative diseases and ischemic injury. The anti-caspase-14 antibody is useful for measuring the level of caspase-14 in a tissue sample. An antibody that binds to a caspase-14 polypeptide is useful for isolating the polypeptide, and an antibody that binds to the large or small subunit the polypeptide is useful for identifying samples with caspase-14 processing activity. An antibody that binds to caspase-14 heterodimer or heterotetramer is useful for isolating caspase-14 with apoptotic activity or in a screening assay to identify (ant)agonists. The antibodies form kits for such purposes. The anti-caspase-14 antibody is also useful for preparing anti-idiotypic antibodies, which mimic a caspase-14 epitope recognized by the anti-caspase-14 antibody. Therefore, the antibody is useful as a competitor of caspase-14 in reducing the level of caspase-14 activity, which reduces the level of apoptotic activity. Oligonucleotides made from the polynucleotides are useful as polymerase chain reaction (PCR) primers or probes to screen genomic or cDNA libraries for similar caspase-14 encoding polynucleotides, or for diagnosis of diseases associated with enhanced or inhibited apoptosis. The isolated caspase-14 gene permits methods of modulating apoptosis for the treatment of human diseases.

CC Sequence 281 AA;

Query Match 24.5%; Score 63; DB 20; Length 281;  
Best Local Similarity 42.3%; Pred. No. 1.1;  
Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 14 GYRSYQHDLRAYGFWRLLVRRFVHRR 39  
Db 208 gylsyrndeksgyfigltvdfihkk 233

RESULT 7  
AA93213  
ID AAY93213 standard; Protein; 281 AA.

AC AAY93213;

DT 04-SEP-2000 (first entry)

DE Amino acid sequence of a murine caspase-14.

KW Caspase-14; cell death specific protease; apoptosis stimulator;  
KW apoptosis; AIDS; neurodegenerative disease; ischemic injury.

OS Mus sp.

FH Key Location/Qualifiers

FT Region 1..156  
FT Active-site /note="large subunit"  
FT Cleavage-site 156..157  
FT Region 163..257  
FT /note="small subunit"

PN WO200028047-A1.

PD 18-MAY-2000.

PE 29-OCT-1999; 99WO-US25523.

PR 06-NOV-1998; 98US-0187789.

PA (UYJE-) UNIV JEFFERSON THOMAS.

PI Alnemri ES, Fernandez-Alnemri T;

DR WPI; 2000-376558/32.

DR N-PSDB; AAA15163.

PT Novel nucleic acids encoding cell death specific protease termed  
PT caspase-14 useful for treating cancers by stimulating apoptosis -  
PS Claim 52; Fig 1; 78pp; English.

CC The present sequence represents a murine caspase-14 polypeptide. The  
CC polypeptide is a cell death specific protease, and is an apoptosis  
CC stimulator. Caspase-14 polynucleotides and polypeptides, and  
CC anti-caspase-14 antibodies are useful for treating or reducing the  
CC severity of pathological conditions associated with increased or  
CC decreased levels of apoptosis. Apoptosis mediated diseases such as  
CC AIDS, neurodegenerative diseases and ischemic injury are treated by  
CC administering anti-caspase-14 antibodies. The antibody is useful for  
CC determining the presence or the level of caspase-14 in tissue sample  
CC and also for the isolation of caspase-14 with apoptotic activity or  
CC in screening assay to identify an agent that inhibits heterodimer or  
CC heterotetramer formation and therefore, apoptosis.

SQ Sequence 281 AA;

Query Match 24.5%; Score 63; DB 21; Length 281;  
Best Local Similarity 42.3%; Pred. No. 1.1;  
Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 14 GYRSYQHDLRAYGFWRLLVRRFVHRR 39  
Db 208 gylsyrndeksgyfigltvdfihkk 233

RESULT 8  
AAW78169  
ID AAW78169 standard; Protein; 398 AA.

AC AAW78169;

DT 13-APR-1999 (first entry)

DE Human secreted protein encoded by gene 44 clone HEPFJ05.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

FH Key Location/Qualifiers  
FT Misc-difference 13  
FT /label= unknown  
FT Misc-difference 258  
FT /label= unknown  
FT Misc-difference 398  
FT /label= unknown

PN WO9856804-A1.

PD 17-DEC-1998.

PE 11-JUN-1998; 98WO-US12125.

PR 02-OCT-1997; 97US-0061060.

PR 13-JUN-1997; 97US-0049547.

PR 13-JUN-1997; 97US-0049548.

PR 13-JUN-1997; 97US-0049549.

PR 13-JUN-1997; 97US-0049550.

PR 13-JUN-1997; 97US-0049606.

PR 13-JUN-1997; 97US-0049607.

PR 13-JUN-1997; 97US-0049608.

PR 13-JUN-1997; 97US-0049609.  
 PR 13-JUN-1997; 97US-0049610.  
 PR 13-JUN-1997; 97US-0049611.  
 PR 13-JUN-1997; 97US-0050566.  
 PR 13-JUN-1997; 97US-0050901.  
 PR 13-JUN-1997; 97US-0052989.  
 PR 08-JUL-1997; 97US-0051919.  
 PR 18-AUG-1997; 97US-0055984.  
 PR 12-SEP-1997; 97US-0058665.  
 PR 12-SEP-1997; 97US-0058668.  
 PR 12-SEP-1997; 97US-0058669.  
 PR 12-SEP-1997; 97US-0058750.  
 PR 12-SEP-1997; 97US-0058971.  
 PR 12-SEP-1997; 97US-0058972.  
 PR 12-SEP-1997; 97US-0058975.  
 PR 02-OCT-1997; 97US-0060834.  
 PR 02-OCT-1997; 97US-0060841.  
 PR 02-OCT-1997; 97US-0060844.  
 PR 02-OCT-1997; 97US-0060865.  
 PR 02-OCT-1997; 97US-0061059.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;  
 PI Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P,  
 PI Yu GL;  
 XX  
 DR WPI: 1999-080881/07.  
 DR N-PSDB; AAX04354.  
 XX  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS  
 XX  
 PS Claim 11: Page 288-289; 380pp; English.  
 CC  
 CC This sequence represents a secreted human protein encoded by the gene  
 CC clone detailed in the descriptor line. The gene can be used to generate  
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc  
 CC portion (e.g. AAX04302) for increasing the stability of the fused  
 CC protein as compared to the human protein only.  
 CC The invention relates to 86 novel genes and their fragments (nucleic  
 CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological  
 CC conditions can be diagnosed by determining the amount of the new  
 CC polypeptides in a sample or by determining the presence of mutations in  
 CC the new polynucleotides. Specific uses are described for each of the 86  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAX04311 for described uses).  
 XX  
 SO Sequence 398 AA;  
 Query Match 22.8%; Score 58.5; DB 20; Length 398;  
 Best Local Similarity 40.0%; Pred. No. 6.8;  
 Matches 12; Conservative 5; Mismatches 8; Indels 5; Gaps 1;  
 QY 6 PWTAMAFHGYSYQHDLRAYG-----FWRL 30  
 Db 36 pgsawawpfgfqlqeqiraagalskrywL 65  
 RESULT 9  
 AAU14277 9  
 ID AAU14277 standard; Protein: 404 AA.  
 XX  
 AC AAU14277;  
 XX  
 XX 24-OCT-2001 (first entry)  
 XX Human novel protein #148.  
 DE  
 XX

KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;  
 KW immunomodulatory; cytosolic; neuroprotective; vulnerrary; nootropic;  
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
 KW antibacterial; antiallergic; dermatological; haemostatic; antistatic;  
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KW tissue regeneration; immune disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20015437-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PE 25-JAN-2001; 2001WO-US02623.  
 XX  
 PR 25-JAN-2000; 2000US-0491404.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-451939/48.  
 DR N-PSDB; AAS22582.  
 XX  
 PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
 PT nervous system disorders, and for regenerating bone and cartilage -  
 PS  
 XX  
 XX Example 4: Page 606-607; 894pp; English.  
 XX  
 CC The invention relates to polynucleotides encoding novel human  
 CC proteins or their active domains. The polypeptides, polynucleotides and  
 CC antibodies raised against the polypeptides are used in a method of  
 CC treatment of a mammal and prevention of disorders caused by the aberrant  
 CC protein expression or activity. The polypeptides can be used as  
 CC molecular weight markers, food supplements, and in antibody production.  
 CC The polypeptides are used to identify compounds which bind to the  
 CC polypeptides. Polynucleotides of the invention are used as probes and  
 CC primers, for sequencing, for chromosome or gene mapping, in the  
 CC production of recombinant proteins, and in generating anti-sense DNA or  
 CC RNA and in gene therapy. Polypeptides of the invention can be used to  
 CC target drugs to a tumour, in assays to determine biological activity, to  
 CC raise antibodies/elicits an immune response, to determine quantitative  
 CC protein levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
 CC anti-inflammatory diseases, nervous system disorders, and infection.  
 CC The present sequence represents a protein of the invention.  
 XX  
 SO Sequence 404 AA;  
 Query Match 22.8%; Score 58.5; DB 22; Length 404;  
 Best Local Similarity 40.0%; Pred. No. 7;  
 Matches 12; Conservative 5; Mismatches 8; Indels 5; Gaps 1;  
 QY 6 PWTAMAFHGYSYQHDLRAYG-----FWRL 30  
 Db 43 pgsawawpfgfqlqeqiraagalskrywL 72  
 RESULT 10  
 AAY75223 10  
 ID AAY75223 standard; Protein: 459 AA.  
 XX  
 AC AAY75223;  
 XX

```

XX 21-MAR-2000 (first entry)
DE Neisseria gonorrhoeae ORF 607 protein sequence SEQ ID NO:1920.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN W09957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizze M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR N-PSDB: AA53985.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2: Page 967; 1453pp; English.
XX
XX AA53015 to AA54536, AA54577 to AA54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA54537 to AA54576 and AA54616 to AA5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 459 AA:

```

Query Match 22.0%; Score 56.5; DB 21; Length 459;  
 Best Local Similarity 32.5%; Pred. No. 15;  
 Matches 13; Conservative 6; Mismatches 12; Indels 9; Gaps 2;

QY 2 VSRVP-----WTAMAFHGR-RSYQHDLRATGFWRLV 32  
 I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I  
 Db 392 vtkvpmfhaaafwgcgllpgyllayfdmglygfwtail 431

RESULT 11  
 AAY75225  
 ID AAY75225 standard; Protein; 459 AA.  
 XX AAY75225;  
 AC  
 XX

```

DT 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 607 protein sequence SEQ ID NO:1924.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN W09957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizze M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR N-PSDB: AA53987.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2: Page 969; 1453pp; English.
XX
XX AA53015 to AA54536, AA54577 to AA54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA54537 to AA54576 and AA54616 to AA5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 459 AA:

```

Query Match 22.0%; Score 56.5; DB 21; Length 459;  
 Best Local Similarity 32.5%; Pred. No. 15;  
 Matches 13; Conservative 6; Mismatches 12; Indels 9; Gaps 2;

QY 2 VSRVP-----WTAMAFHGR-RSYQHDLRATGFWRLV 32  
 I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I I:::I  
 Db 392 vtkvpmfhaaafwgcgllpgyllayfdmglygfwtail 431

RESULT 12  
 AAY93216  
 ID AAY93216 standard; Protein; 214 AA.  
 XX AAY93216;  
 AC  
 XX 04-SEP-2000 (first entry)

XX	Amino acid sequence of a human caspase-14 splice variant.
DE	
XX	Caspase-14; cell death specific protease; apoptosis stimulator;
KW	apoptosis; AIDS; neurodegenerative disease; ischemic injury.
XX	
OS	Homo sapiens.
XX	
FH	Location/Qualifiers
FT	Active-site 102..106
XX	
PN	MO200028047-A1.
PD	18-MAY-2000.
XX	
PE	29-OCT-1999; 99WO-US25523.
XX	
PR	06-NOV-1998; 98US-0187789.
XX	
PA	(UYJE-) UNIV JEFFERSON THOMAS.
XX	
PI	Alnemri ES, Fernandez-Alnemri T;
XX	
DR	WPI: 2000-376558/32.
DR	N-PsDB: AAA15166.
XX	
PT	Novel nucleic acids encoding cell death specific protease termed
PS	caspase-14 useful for treating cancers by stimulating apoptosis -
XX	
PS	Claim 42; Fig 9; 78pp; English.
XX	
CC	The present sequence represents a human caspase-14 splice variant. The
CC	polypeptide is a cell death specific protease, and is an apoptosis
CC	stimulator. Caspase-14 polynucleotides and polypeptides, and
CC	anti-caspase-14 antibodies are useful for treating or reducing the
CC	severity of pathological conditions associated with increased or
CC	decreased levels of apoptosis. Apoptosis mediated diseases such as
CC	AIDS, neurodegenerative diseases and ischemic injury are treated by
CC	administering anti-caspase-14 antibodies. The antibody is useful for
CC	determining the presence or the level of caspase-14 in tissue sample
CC	and also for the isolation of caspase-14 with apoptotic activity or
CC	in screening assay to identify an agent that inhibits heterodimer or
CC	heterotrimer formation and therefore, apoptosis.
XX	
SQ	Sequence ~ 214 AA;
XX	
Query Match	21.8%; Score 56; DB 21; Length 214;
Best Local Similarity	31.8%; Pred. No. 7.6;
Matches 14; Conservative 7; Mismatches 17; Indels 6; Gaps 1;	
OY	5 VPMTAMAFH-----GYRSYQHDLRAYGFWRLVRFVRRPHV 42
:	:        :        :      :
Ddb	131 iptycdalhystvtegylayrhdqgscldqlvavfckrkghl 174
RESULT 13	
ID	AAV68865
XX	AAV68865 standard; Protein: 229 AA.
XX	
AC	AAV68865;
XX	
DT	16-MAY-2000 (first entry)
XX	
DE	Amino acid sequence of a human caspase-like polypeptide.
XX	
KW	Mouse; caspase-like polypeptide; human; caspase; apoptosis;
KW	skin disease; keratinisation; wound healing.
XX	
OS	Homo sapiens.
XX	
PN	MO200004169-A1.
XX	

```

PD 27-JAN-2000.
PF 12-JUL-1999; 99WO-EP04939.
XX 17-JUL-1998; 98EP-0202422.
XX (VLAAS-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX Van De Craen M, Declercq W, Vandenabeele P, Fiers W;
DR WPI: 2000-182433/16.
DR N-PSDB; AAZ60684.
XX
XX New murine and human caspase homologues useful for treating skin
PT related disorders -
XX
PS Claim 2; Page 53-54; 68pp; English.
XX
XX The present sequence represents a human caspase-1like polypeptide.
CC The specification also describes a murine caspase-1like polypeptide.
CC Caspases are cysteine/hl aspartate-specific proteinases which play a
CC central role in apoptosis. The polypeptides of the invention are related
CC to human and murine caspase-2 and human caspase-9, and possess all of
CC the typical amino acids involved in catalysis, including the QACRG box,
CC and contain no or only a very short prodomain. mRNA expression of the
CC homologues of the invention is predominant in the skin. The caspase-1like
CC polypeptides are useful for treating human or animal diseases, such
CC as skin diseases. They are also useful for screening for compounds that
CC modulate its activity, i.e. agonists, antagonists, and inhibitors. The
CC caspase-1like polypeptides and polynucleotides are useful for modulating
CC keratinisation, for diagnosing and treating inappropriate wound
CC healing.
XX
SQ Sequence 229 AA;
Query Match 21.8%; Score 56; DB 21; Length 229;
Best Local Similarity 31.8%; Pred. No. 8.2;
Matches 14; Conservative 7; Mismatches 17; Indels 6; Gaps 1;
QY 5 VPEWTAMAFH-----GYRSYQHDLPAYGFWRLVRFVHRPQV 42
Db 146 iptycdalhwystvegyiaiyndqkysctqlclvdvfkirkghi 189
RESULT 14
AAV93214
ID AAV93214 standard; Protein; 242 AA.
XX
XX AAV93214;
XX
XX 04-SEP-2000 (first entry)
XX
XX Amino acid sequence of a human caspase-14.
XX
XX Caspase-14; cell death specific protease; apoptosis stimulator;
KW apoptosis; AIDS; neurodegenerative disease; ischemic injury.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 1..146
XX Active-site 130..134 /note= "large subunit"
XX Cleavage-site 146..147
XX Region 147..242
XX /note= "small subunit"
XX
XX WO2000028047-A1.
XX
XX 18-MAY-2000.
XX
XX 29-OCT-1999; 99WO-US25523.

```

```

xx PR      06-NOV-1998;       98US-0187789.
xx XX
xx PA      (UYJE-) UNIV JEFFERSON THOMAS.
xx PI
xx ALnemri ES, Fernandez-Alnemri T;
xx DR      WPI: 2000-376558/32.
xx N-PSDB; AAA15164.
xx PT      Novel nucleic acids encoding cell death specific protease termed
xx caspase-14 useful for treating cancers by stimulating apoptosis
xx PS      Claim 13; Fig 7; 78pp; English.
xx XX
xx CC      The present sequence represents a human caspase-14 polypeptide. The
xx polypeptide is a cell death specific protease, and is an apoptosis
xx stimulator. Caspase-14 polynucleotides and polypeptides, and
xx anti-caspase-14 antibodies are useful for treating or reducing the
xx severity of pathological conditions associated with increased or
xx decreased levels of apoptosis. Apoptosis mediated diseases such as
xx AIDS, neurodegenerative diseases and ischemic injury are treated by
xx administering anti-caspase-14 antibodies. The antibody is useful for
xx determining the presence or the level of caspase-14 in tissue sample
xx and also for the isolation of caspase-14 with apoptotic activity or
xx in screening assay to identify an agent that inhibits heterodimer or
xx heterotrimer formation and therefore, apoptosis.
xx SQ      Sequence      242 AA;

Query Match          21.8%; Score 56; DB 21; Length 242;
Best Local Similarity 31.8%; Pred. No. 8.7;
Matches   14; Conservative    7; Mismatches   17; Indels     6; Gaps     1.

Oy      5 VEWTAFAFH-----GYREYOHDLRAYGFWRLLVREFVRRPHV 42
         | | | | | | | | | | | | | | | | | | | | | |
Db      159 iptydahvystvegyiaiyndqkysctfqlvdvftkrkghi 202

RESULT 15
AAAG91100
ID      AAG91100 standard; Protein; 731 AA.
AC      AAG91100;
AD
AE
AF      26-SEP-2001 (first entry)
AG
AH
AI
AJ
AK
AL
AM      C glutamicum protein fragment SEQ ID NO: 4854.
AN
AO
AP
AQ
AR      Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
AS      organic acid synthesis.
AT
AU      Corynebacterium glutamicum.
AV
AW      EPI108790-A2.
AX
AY      20-JUN-2001.
AZ
BA      18-DEC-2000; 2000EP-0127688.
BB
BC      16-DEC-1999; 99JP-0377484.
BD      07-APR-2000; 2000JP-0159162.
BE      03-AUG-2000; 2000JP-0280988.
BF
BG      (KYOW ) KYOMA HAKKO KOGYO KK.
BH
BI      Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
BJ      Tateishi N, Senoh A, Ikeda M, Ozaki A;
BK      WPI: 2001-376931/40.
BL      N-PSDB; AAH66319.
BM
BN
BO
BP
BQ
BR
BS
BT
BU
BV
BW
BX
BY
BZ
CA
CB
CC
CD
CE
CF
CG
CH
CI
CJ
CK
CL
CM
CN
CO
CP
CQ
CR
CS
CT
CU
CV
CW
CX
CY
CZ
DA
DB
DC
DD
DE
DF
DG
DH
DI
DJ
DK
DL
DM
DN
DO
DP
DQ
DR
DS
DT
DU
DV
DW
DX
DY
EZ

```

PT Novel polynucleotides derived from *Corynebacterium* bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 PS  
 PS Claim 17; SEQ ID NO: 4854; 246bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the *Corynebacterium* bacterium *Corynebacterium* glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of *Corynebacterium* bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from *Corynebacterium* bacterium, and identifying a homologue of a gene derived  
 CC from *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 XX  
 SQ Sequence 731 AA:

	Query Match	21.8%	Score 56;	DB 22;	Length 731;
	Best Local Similarity	28.3%	Pred. No.	30;	
	Matches 13; Conservative	7;	Mismatches 12;	Indels 14;	Gaps 2
Oy	1 TVSRPWTAWA-----FHGRSTQHDLRAY---GFNRLLY 32				
	: :				
Db	140 tvrgatcfaapnaigacavagifmgwnaaqdpmrirngasgllveilt 185				

Search completed: August 15, 2002, 11:59:15  
Job time: 530 sec



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 11:55:35 ; Search time 20.64 Seconds  
(without alignments)  
53.253 Million cell updates/sec

Title: DEVI-613-COMB

Perfect score: 257  
Sequence: 1 TVSRVPMWTAMAFHGYSYOH.....GFWRLVRRFVRRPHVESQ 45

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/plodata/2/1aa/5A-COMB.pep.\*  
2: /cgn2\_6/plodata/2/1aa/5B-COMB.pep.\*  
3: /cgn2\_6/plodata/2/1aa/6A-COMB.pep.\*  
4: /cgn2\_6/plodata/2/1aa/6B-COMB.pep.\*  
5: /cgn2\_6/plodata/2/1aa/PCUTUS-COMB.pep.\*  
6: /cgn2\_6/plodata/2/1aa/Backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	24.5	53	4	US-09-187-789-65
2	63	24.5	260	4	US-09-187-789-2
3	57	22.2	144	1	US-07-956-700B-89
4	57	22.2	144	1	US-08-476-537-89
5	57	22.2	144	1	US-08-485-607-89
6	56	22.2	144	2	US-08-475-879-89
7	56	21.8	214	4	US-09-187-789-9
8	56	21.8	242	4	US-09-187-789-5
9	53	20.6	215	2	US-08-729-152-1
10	52	20.2	66	4	US-08-936-165A-376
11	52	20.2	244	2	US-08-933-750C-23
12	52	20.2	244	4	US-09-234-613-23
13	51	19.8	84	4	US-09-247-155-173
14	50.5	19.6	144	1	US-07-956-700B-79
15	50.5	19.6	144	1	US-08-476-537-89
16	50.5	19.6	144	1	US-08-485-607-79
17	50.5	19.6	144	2	US-08-475-879-79
18	50.5	19.6	599	3	US-08-556-419-22
19	50.5	19.6	629	3	US-08-556-419-23
20	49.5	19.3	495	6	5516630-4
21	49.5	19.3	1463	4	US-08-157-005-3
22	49.5	19.3	1463	4	US-08-747-863-3
23	49	19.1	361	3	US-08-652-265-22
24	49	19.1	361	4	US-08-834-497A-22
25	49	19.1	361	4	US-09-503-444A-22
26	48.5	18.9	2227	3	US-08-475-886-2
27	48.5	18.9	2227	3	US-08-475-886-4

28	48.5	18.9	2227	3	US-08-475-886-6	Sequence 6, Appl1
29	48.5	18.9	2227	4	US-08-397-232-2	Sequence 2, Appl1
30	48.5	18.9	2227	4	US-08-397-232-4	Sequence 2, Appl1
31	48.5	18.9	2227	4	US-09-171-387-2	Sequence 2, Appl1
32	47.5	18.5	997	3	US-08-872-094-8	Sequence 8, Appl1
33	47	18.3	255	2	US-08-685-992-16	Sequence 16, Appl1
34	47	18.3	255	2	US-09-144-925-16	Sequence 2, Appl1
35	47	18.3	276	1	US-07-882-339-2	Sequence 2, Appl1
36	47	18.3	276	1	US-08-183-213-2	Sequence 2, Appl1
37	46.5	18.1	4150	3	US-09-428-517-2	Sequence 2, Appl1
38	46	17.9	148	3	US-08-946-329A-65	Sequence 65, Appl1
39	46	17.9	466	4	US-08-868-435-31	Sequence 31, Appl1
40	46	17.9	466	4	US-08-868-435-31	Sequence 31, Appl1
41	46	17.9	573	6	5215909-12	Patent No. 5215909
42	46	17.9	635	6	5215909-10	Patent No. 5215909
43	45.5	17.7	444	1	US-09-178-002-2	Sequence 2, Appl1
44	45.5	17.7	466	3	US-08-704-711A-17	Sequence 17, Appl1
45	45.5	17.7	467	1	US-09-178-002-4	Sequence 4, Appl1

## ALIGNMENTS

RESULT 1  
US-09-187-789-65  
; Sequence 65, Application US/09187789  
; Patent No. 6340740  
; GENERAL INFORMATION:  
; APPLICANT: Alnemrl, Emed S.  
; APPLICANT: Fernandez-Alnemrl, Teresa  
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
; FILE REFERENCE: 480140.434C1  
; CURRENT APPLICATION NUMBER: US/09/187,789  
; CURRENT FILING DATE: 1998-11-06  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 65  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-187-789-65

Query Match 24.5%; Score 63; DB 4; Length 53;

Best Local Similarity 42.3%; Pred. No. 0.06; Mismatches 9; Indels 0; Gaps 0;

Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 14 GYSYQHDURAYGFWRLVRRFVRR 39

DB 28 GYLSYRHDEKSGFTQTLTDVPIHKK 53

RESULT 2  
US-09-187-789-2  
; Sequence 2, Application US/09187789  
; Patent No. 6340740  
; GENERAL INFORMATION:  
; APPLICANT: Alnemrl, Emed S.  
; APPLICANT: Fernandez-Alnemrl, Teresa  
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
; FILE REFERENCE: 480140.434C1  
; CURRENT APPLICATION NUMBER: US/09/187,789  
; CURRENT FILING DATE: 1998-11-06  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-187-789-2

Query Match 24.5%; Score 63; DB 4; Length 260;  
Best Local Similarity 42.3%; Pred. No. 0.35;  
Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 14 GYRSHDRLRAYGFWRLVRRFVRR 39  
Db 192 GYLSTHRDGRSGFTOTLDTVFTHK 217

## RESULT 3

US-07-956-700B-89  
; Sequence 89, Application US/07956700B  
; Patent No. 5539092  
; GENERAL INFORMATION:  
; APPLICANT: Robert Haselkorn and Piotr Gornicki  
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
; TITLE OF INVENTION: Carboxylase  
; NUMBER OF SEQUENCES: 116  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: 321 No. 5539092th Clark Street  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/956,700B  
; FILING DATE: 19921002  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Thomas E. No. 5539092thrup  
; REGISTRATION NUMBER: 33,268  
; REFERENCE/DOCKET NUMBER: ARCD:058  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-312-744-0090  
; TELEFAX: 1-312-755-4489  
; INFORMATION FOR SEQ. ID NO: 89:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 144 amino acids  
; TYPE: Amino acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; MOLECULE TYPE: Peptide  
US-07-956-700B-89

Query Match 22.2%; Score 57; DB 1; Length 144;  
Best Local Similarity 34.1%; Pred. No. 1.2;  
Matches 15; Conservative 8; Mismatches 19; Indels 2; Gaps 2;

QY 4 RVPM-TAMAFHGR-SYQHDRLRAYGFWRLVRRFVRRPVESQ 45  
Db 32 RIMWDETRDGRFSSQEAASSFGDDRLLEKFIIDNPRIEIO 75

## RESULT 4

US-08-476-537-89  
; Sequence 89, Application US/08476537  
; Patent No. 5756290  
; GENERAL INFORMATION:  
; APPLICANT: Robert Haselkorn and Piotr Gornicki  
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
; NUMBER OF SEQUENCES: 116  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: 321 No. 5756290th Clark Street

CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,537  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/956,700  
FILING DATE: 10/21/92

ATTORNEY/AGENT INFORMATION:

NAME: Thomas E. No. 5756290thrup  
REGISTRATION NUMBER: 33,268  
REFERENCE/DOCKET NUMBER: ARCD:058  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-312-744-0090  
TELEFAX: 1-312-755-4489

INFORMATION FOR SEQ. ID NO: 89:

SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-476-537-89

Query Match 22.2%; Score 57; DB 1; Length 144;  
Best Local Similarity 34.1%; Pred. No. 1.2;  
Matches 15; Conservative 8; Mismatches 19; Indels 2; Gaps 2;

QY 4 RVPM-TAMAFHGR-SYQHDRLRAYGFWRLVRRFVRRPVESQ 45  
Db 32 RIMWDETRDGRFSSQEAASSFGDDRLLEKFIIDNPRIEIO 75

## RESULT 5

US-08-485-607-89  
; Sequence 89, Application US/08485607  
; Patent No. 5792627  
; GENERAL INFORMATION:  
; APPLICANT: Robert Haselkorn and Piotr Gornicki  
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
; NUMBER OF SEQUENCES: 116  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: 321 No. 5792627th Clark Street  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,607  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/956,700  
; FILING DATE: 10/21/92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Thomas E. No. 5792627thrup  
; REGISTRATION NUMBER: 33,268

REFERENCE/DOCKET NUMBER: ARCD:058  
TELEPHONE: 1-312-744-0090  
TELEFAX: 1-312-755-4489  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-485-607-89

Query Match 22.2% Score 57; DB 1; Length 144;  
Best Local Similarity 34.1%; Pred. No. 1.2;  
Matches 15: Conservative 8; Mismatches 19; Indels 2; Gaps 2;

OY 4 RVPW-TAMAFHGYR-SYOHDLRAYGFWRLLVRRFVRRPVSQ 45  
Db 32 RIPMDEETRDGFRSSQEAASSFGDDRLLEKFIIDNRHIEIQ 75

RESULT 6  
US-08-475-879-89  
Sequence 89, Application US/08475879  
Patent No. 5972644  
Patent No. 5972644 5786170  
GENERAL INFORMATION:  
APPLICANT: Robert Haselkorn and Piotr Gornicki  
TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
NUMBER OF SEQUENCES: 116  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: 321 No. 5972644 5786170th Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,879  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/956,700  
FILING DATE: 10/21/92  
ATTORNEY/AGENT INFORMATION:  
NAME: Thomas E. NO. 5972644 5786170thrup  
REGISTRATION NUMBER: 33,268  
REFERENCE/DOCKET NUMBER: ARCD:058  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-312-744-0090  
TELEFAX: 1-312-755-4489  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-475-879-89

Query Match 22.2% Score 57; DB 2; Length 144;  
Best Local Similarity 34.1%; Pred. No. 1.2;  
Matches 15: Conservative 8; Mismatches 19; Indels 2; Gaps 2;

OY 4 RVPW-TAMAFHGYR-SYOHDLRAYGFWRLLVRRFVRRPVSQ 45  
Db 32 RIPMDEETRDGFRSSQEAASSFGDDRLLEKFIIDNRHIEIQ 75

RESULT 7  
US-09-187-789-9  
Sequence 9, Application US/09187789  
Patent No. 6340740  
GENERAL INFORMATION:  
APPLICANT: Alnemrl, Emdad S.  
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
FILE REFERENCE: 480140.434C1  
CURRENT APPLICATION NUMBER: US/09/187,789  
CURRENT FILING DATE: 1998-11-06  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 9  
LENGTH: 214  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-187-789-9

Query Match 21.8% Score 56; DB 4; Length 214;  
Best Local Similarity 31.8%; Pred. No. 2.5;  
Matches 14: Conservative 7; Mismatches 17; Indels 6; Gaps 1;

OY 5 VPMWTAMAFH-----GYRSYOHDLRAYGFWRLLVRRFVRRPVSQ 42  
Db 131 IPTYTDALHYSTVGITAYARRHDKGSCFIOTLVDFVFKRKGHI 174

RESULT 8  
US-09-187-789-5  
Sequence 5, Application US/09187789  
Patent No. 6340740  
GENERAL INFORMATION:  
APPLICANT: Alnemrl, Emdad S.  
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
FILE REFERENCE: 480140.434C1  
CURRENT APPLICATION NUMBER: US/09/187,789  
CURRENT FILING DATE: 1998-11-06  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 5  
LENGTH: 242  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-187-789-5

Query Match 21.8% Score 56; DB 4; Length 242;  
Best Local Similarity 31.8%; Pred. No. 2.9;  
Matches 14: Conservative 7; Mismatches 17; Indels 6; Gaps 1;

OY 5 VPMWTAMAFH-----GYRSYOHDLRAYGFWRLLVRRFVRRPVSQ 42  
Db 159 IPTYTDALHYSTVGITAYARRHDKGSCFIOTLVDFVFKRKGHI 202

RESULT 9  
US-08-729-152-1  
Sequence 1, Application US/08729152  
Patent No. 5871739  
GENERAL INFORMATION:  
APPLICANT: Inoue, Eri  
TITLE OF INVENTION: Pharmaceutical Composition  
NUMBER OF SEQUENCES: 47

```

CORRESPONDENCE ADDRESS:
ADDRESS: Wenderoth, Lind & Ponack
STREET: Southern Building, Suite 700, 805 Fifteenth
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE:

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,152
FILING DATE: 11-OCT-1996
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 291993/1995
FILING DATE: 13-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mairen M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:

TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 215 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: protein
US-08-729-152-1

Query Match      20.6%; Score 53; DB 2; Length 215;
Best Local Similarity 34.5%; Pred. No. 6.5;
Matches 10; Conservative 4; Mismatches 9; Indels 6; Gaps 1;

QY 7 WTMAFHGRSYOHDIRA-----YGFWR 29
Db 168 WVCYHVLGRGYOYLKCDHHEGDYRHR 196

RESULT 10
US-08-936-165A-376
Sequence 376, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Richard
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: NO. 6348582el Prokaryotic Polynucleotides,
TITLE OF INVENTION: Polypeptides and Their Uses
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,165A
FILING DATE: 24-SEP-1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/027,032
FILING DATE: 24-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm1, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 376:
SEQUENCE CHARACTERISTICS:
LENGTH: 66 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-936-165A-376

Query Match      20.2%; Score 52; DB 4; Length 66;
Best Local Similarity 32.4%; Pred. No. 2.4;
Matches 11; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 17 SYOHDLRAVGFWRLLVRR-----FVHRRPHVS 44
Db 25 SYEPIACGTWRVYQRLKALEVLTHERVHEN 58

RESULT 11
US-08-933-750C-23
Sequence 23, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
```

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 244 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SPLNOT04  
CLONE: 1561587  
US-08-933-750C-23

Query Match 20.2%; Score 52; DB 2; Length 244;  
Best Local Similarity 75.0%; Pred. No. 10;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 32 VRFFVHRPHVE 43  
|||:||||:  
Db 104 VRFFVHRPHVD 115

RESULT 12  
US-09-234-613-23  
Sequence 23, Application US/09234613  
Patent No. 6132973

GENERAL INFORMATION:

APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Bandman, Olga  
APPLICANT: Shah, Purvi  
APPLICANT: Au-Young, Janice  
APPLICANT: Yue, Henry  
APPLICANT: Guebler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/234, 613

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750

FILING DATE: September 23, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0356 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:

LENGTH: 244 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SPLNOT04  
CLONE: 1561587  
US-09-234-613-23

Query Match 20.2%; Score 52; DB 4; Length 244;  
Best Local Similarity 75.0%; Pred. No. 10;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 32 VRFFVHRPHVE 43  
|||:||||:  
Db 104 VRFFVHRPHVD 115

RESULT 13

US-09-247-155-173

Sequence 173, Application US/09247155A

Patent No. 6312922

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
APPLICANT: Duclert, Aymeric  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: Complementary DNAs

FILE REFERENCE: GENSET.021A

CURRENT APPLICATION NUMBER: US/09/247,155A

EARLIER FILING DATE: 1999-02-09

EARLIER APPLICATION NUMBER: 60/074,121

EARLIER FILING DATE: 1998-02-09

EARLIER APPLICATION NUMBER: 60/081,563

EARLIER FILING DATE: 1998-04-13

EARLIER APPLICATION NUMBER: 60/096,116

EARLIER FILING DATE: 1998-08-10

EARLIER APPLICATION NUMBER: 60/099,273

EARLIER FILING DATE: 1998-10-04

NUMBER OF SEQ ID NOS: 182

SOFTWARE: Patent.pm

SEQ ID NO 173

LENGTH: 84

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SIGNAL

LOCATION: -36..-1

FEATURE:

NAME/KEY: UNSURE

LOCATION: -26..-24

OTHER INFORMATION: Xaa = any one of the twenty amino acids

US-09-247-155-173

Query Match 19.8%; Score 51; DB 4; Length 84;

Best Local Similarity 32.6%; Pred. No. 4.3;

Matches 14; Conservative 4; Mismatches 17; Indels 8; Gaps 1;

QY 5 VPMTAMAFHGYSYOHDLRAYGFWR-----LLVRFVHR 39  
|||:||||:  
Db 27 VSMACGCHGAGTQSPGVAGPWRPRPCVGSLLAARSLHKQ 69

RESULT 14

US-07-956-700B-79

Sequence 79, Application US/07956700B

Patent No. 553092

GENERAL INFORMATION:

APPLICANT: Robert Haselkorn and Piotr Gornicki

TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA

TITLE OF INVENTION: Carboxylase

NUMBER OF SEQUENCES: 116

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5539092th Clark Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII-DOS
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/956,700B
; FILING DATE: 19921002
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas E. No. 5539092chrup
; REGISTRATION NUMBER: 33,268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-312-744-0090
; TELEFAX: 1-312-755-4489
;
; INFORMATION FOR SEQ. ID NO: 79:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Peptide
;
; US-07-956-700B-79

Query Match          19.6%; Score 50.5; DB 1; Length 144;
Best Local Similarity 36.4%; Pred. No. 9.2;
Matches 12; Conservative 7; Mismatches 13; Indels 1; Gaps 1;

QY 14 GYR-STQHDIRAYGFWRLLVRRFVHRRPVESQ 45
      |:| | | | | | | | | | | | | | | |
DB 43 GFRLSSQEAASSFGDDRLLEKFIIDNPRIHEIQ 75

RESULT 15
US-08-476-537-79
; Sequence 79, Application US/08476537
; Patent No. 5756290
;
; GENERAL INFORMATION:
; APPLICANT: Robert Haselkorn and Piotr Gornicki
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA
; TITLE OF INVENTION: Carboxylase
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5756290th Clark Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII-DOS
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,537
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/956,700
; FILING DATE: 10/21/92
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas E. No. 5756290chrup
; REGISTRATION NUMBER: 33,268
```

```
;
; REFERENCE/DOCKET NUMBER: ARCD:058
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-312-744-0090
; TELEFAX: 1-312-755-4489
;
; INFORMATION FOR SEQ. ID NO: 79:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Peptide
;
; US-08-476-537-79

Query Match          19.6%; Score 50.5; DB 1; Length 144;
Best Local Similarity 36.4%; Pred. No. 9.2;
Matches 12; Conservative 7; Mismatches 13; Indels 1; Gaps 1;

QY 14 GYR-STQHDIRAYGFWRLLVRRFVHRRPVESQ 45
      |:| | | | | | | | | | | | | | | |
DB 43 GFRLSSQEAASSFGDDRLLEKFIIDNPRIHEIQ 75
```

Search completed: August 15, 2002, 11:59:42  
Job time: 247 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 11:56:50 ; Search time 25.35 seconds  
(without alignments)  
170.573 Million cell updates/sec

Title: DEVI-613-COMB

Perfect score: 257  
Sequence: 1 TVSRVPMWTAMAFHGYSYOH.....GFWRLVRFVHRPHEVSEQ 45

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: PIR-71:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	66.5	25.9	249	1	CYRMB1
2	61.5	23.9	537	2	T45634
3	61.1	23.7	724	2	T16187
4	60.5	23.5	253	2	S07264
5	59.5	23.2	288	2	T42847
6	59.5	23.0	292	2	T43618
7	59.5	23.0	322	2	C70905
8	58	22.6	1175	2	T20346
9	58	22.6	1178	2	A47255
10	58	22.6	1178	1	JC2460
11	58	22.6	1178	2	JC4391
12	57.5	22.4	215	2	JC1231
13	57	22.2	704	2	A34337
14	57	22.2	771	2	H72410
15	56.5	22.0	459	2	D81950
16	56	21.8	242	2	JC7517
17	56	21.8	278	2	S46681
18	55.5	21.6	232	2	JC7706
19	55.5	21.6	300	2	T37759
20	54.5	21.2	151	2	AC0495
21	54	21.0	198	2	S01608
22	54	21.0	198	2	S55512
23	54	21.0	198	2	S55513
24	54	21.0	215	2	S55514
25	54	21.0	215	2	S55515
26	54	21.0	398	2	B82894
27	54	21.0	976	2	S40697
28	53.5	20.8	291	2	T01241
29	53	20.6	32	2	I46165

30	53	20.6	149	2	C27898	beta-crystallin A3
31	53	20.6	177	2	S10088	beta-crystallin A3
32	53	20.6	197	2	JH0602	beta-crystallin A2
33	53	20.6	215	1	CYHUB3	beta-crystallin A3
34	53	20.6	215	2	I45857	beta-crystallin A3
35	53	20.6	238	1	CYCHB1	beta-crystallin B1
36	53	20.6	285	1	IDECRP	replication initia
37	53	20.6	285	2	T00306	replication-asso
38	53	20.6	285	2	T64780	hypothetical prote
39	53	20.6	705	2	T06682	hypothetical prote
40	52.5	20.4	273	2	B83531	probable epoxide h
41	52.5	20.4	320	2	D84664	epoxide hydrolase
42	52.5	20.4	321	2	C84664	hypothetical prote
43	52.5	20.4	500	2	H96570	probable pyruvate-
44	52.5	20.4	1177	2	AG0284	NADH dehydrogenase
45	52	20.2	118	2	G81992	

## ALIGNMENTS

## RESULT 1

CYRTB1

beta-crystallin B1 precursor - rat

N:Contains: beta B1b crystallin

C:Species: Rattus norvegicus (Norway rat)

C>Date: 27-Nov-1985 #sequence\_revision 27-Nov-1985 #text\_change 22-Jun-1999

C:Accession: A02925; B02925; S33586

R:den Dunnen, J.T.; Moormann, R.J.M.; Schoenmakers, J.G.G.

Biochim. Biophys. Acta 824, 295-303, 1985

A:Title: Rat lens beta-crystallins are internally duplicated and homologous to gamma-

A:Reference number: A90654; MUID:85175137

A:Accession: A02925

A:Molecule type: DNA

A:Residues: 1-50 <DE1>

A:Cross-references: GB:X06377; NID:956014; PIDN:CAA29679.1; PID:956015

A:Accession: B02925

A:Molecule type: mRNA

A:Residues: 34-249 <DE2>

R:David, L.L.; Shearer, T.R.

FEBS Lett. 324, 265-270, 1993

A:Title: beta-crystallins insolubilized by calpain II in vitro contain cleavage sites

A:Reference number: S33586; MUID:94009594

A:Accession: S33586

A:Molecule type: protein

A:Residues: 50-55,233,'X',235-238,'X',240 <DAV>

C:Superfamily: beta-crystallin

C:Keywords: duplication; eye lens

F:12-249/Product: beta-crystallin B1b #status predicted <BBC>

F:56-95/Domain: crystallin repeat <GK1>

F:96-140/Domain: crystallin repeat <GK2>

F:146-187/Domain: crystallin repeat <GK3>

F:188-230/Domain: crystallin repeat <GK4>

Query Match 25.9%; Score 66.5; DB 1; Length 249;  
Best Local Similarity 33.3%; Pred. No. 0.38;  
Matches 18; Conservativity 5; Mismatches 20; Indels 11; Gaps 2;

OY 1 TVSRVPMWTAMAFHGYSYOH-----DIRAYGFWRL-----VRFVHRPHEV 43  
Db 184 TVSGGTWVGQYDPGVRGYLYLEPDRFHWNMENCAFQPMQAVRLRDRQNHQ 237

## RESULT 2

T45634

hypothetical protein F13112.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000

C:Accession: T45634

R:Choisme, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Attiguenave, F.;

submitted to the Protein Sequence Database, November 1999

A:Reference number: 223010

A:Accession: T45634  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-537 <CHO>  
 A:Cross-references: EMBL:AL133292  
 A:Experimental source: cultivar Columbia; BAC clone F13112  
 C:Genetics:  
 A:Map position: 3  
 A:introns: 35/2; 153/3; 229/2; 350/3; 399/3; 445/3; 457/1  
 A:Note: F13112.30

Query Match 23.9%; Score 61.5; DB 2; Length 537;  
 Best Local Similarity 45.5%; Pred. No. 3.6;  
 Matches 15; Conservative 3; Mismatches 6; Indels 9; Gaps 2;

OY 3 SRVPTAMAFHGYSYQHDLRAYGFWR-LLVRR 34  
 DB 363 SAVPMSMAFTGYIA-----GFMSDLLIRR 387

RESULT 3  
 T16187  
 hypothetical protein F27D9.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Nov-1999  
 C:Accession: T16187  
 R:Bentley, D.  
 submitted to the EMBL Data Library, February 1996  
 A:Description: The sequence of C. elegans cosmid F27D9.  
 A:Reference number: Z18473  
 A:Accession: T16187  
 A:Molecule type: DNA  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residues: 1-724 <BEN>  
 A:Cross-references: EMBL:U49829; NID:g1203924; PID:g1203927; PIDN:AAA93384.1; CESP:F27D9  
 C:Genetics:  
 A:Gene: CESP:F27D9.5  
 A:introns: 26/2; 58/1; 124/3; 341/3; 437/3; 672/3  
 C:Superfamily: propionyl-CoA carboxylase alpha chain; biotin carboxylase homology; lipoy

Query Match 23.7%; Score 61; DB 2; Length 724;  
 Best Local Similarity 38.6%; Pred. No. 5.7;  
 Matches 17; Conservative 6; Mismatches 19; Indels 2; Gaps 2;

OY 4 RVPWT-AMAFHGYR-SYQHLRAYGFWRLVRRFVRRPHVESQ 45  
 DB 216 RVAMNDKQAREGRLSKQEAASSFGDDRMVKEFDNPHIEHQ 259

RESULT 4  
 S07264  
 beta-crystallin B1 - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 13-Aug-1999  
 C:Accession: S07264; A27898  
 R:Quax-Jaeken, Y.; Janssen, C.; Quax, W.; van den Heuvel, R.; Bloemendal, H.  
 J. Mol. Biol. 180, 457-472, 1984  
 A:Title: Bovine beta-crystallin complementary DNA clones. Alternating proline / alanine  
 A:Reference number: S07264; MUID:85134866  
 A:Accession: S07264  
 A:Molecule type: mRNA  
 A:Residues: 1-253 <QUA>  
 A:Cross-references: EMBL:X01808; NID:g289; PIDN:CAA25951.1; PID:g290  
 R:Berbers, G.A.M.; Hoekman, W.A.; Bloemendal, H.; de Jong, W.W.; Kleinschmidt, T.; Braun  
 Eur. J. Biochem. 139, 467-479, 1984  
 A:Title: Homology between the primary structures of the major bovine beta-crystallin cha  
 A:Reference number: A91135; MUID:84132067  
 A:Accession: A27898  
 A:Molecule type: protein  
 A:Residues: 2'12',4'-7',9'-10',A',12-50,52-93;103-146,'S',148-253 <BBR>  
 A:Experimental source: lens cortex

C:Superfamily: beta-crystallin  
 C:Keywords: blocked amino end; duplication; eye lens  
 F:26-53/Region: alanine/proline-rich  
 F:2/Modified site: blocked amino end (ser) (in mature form) (probably acetylated) #st

Query Match 23.5%; Score 60.5; DB 2; Length 253;  
 Best Local Similarity 32.1%; Pred. No. 2.3;  
 Matches 17; Conservative 5; Mismatches 20; Indels 11; Gaps 2;

OY 2 VSRVPTAMAFHGYSYQH-----DLRAYGFWRL-----VRRFVRRPHVE 43  
 DB 189 VSSGTWVGIOYFGYRGYQYLLEPGDFRHNWNCAGFOQOMAVRRLDRQWRE 241

RESULT 5  
 T42847  
 replication protein - Yersinia pestis plasmid pCD1  
 C:Species: Yersinia pestis  
 C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 21-Jul-2000  
 C:Accession: T42847  
 R:Perry, R.D.; Straley, S.C.; Fetherston, J.D.; Rose, D.J.; Gregor, J.; Blattner, F.R  
 Infect. Immun. 66, 4611-4623, 1998  
 A:Title: DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of Yersini  
 A:Reference number: Z22273; MUID:98427122  
 A:Accession: T42847  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-288 <PER>  
 A:Cross-references: EMBL:AF074612; NID:g3822037; PIDN:AAC69762.1; PID:g3822042  
 A:Experimental source: strain KIM5  
 C:Genetics:  
 A:Genome: plasmid pCD1  
 A:Note: repA  
 C:Superfamily: repl protein

Query Match 23.2%; Score 59.5; DB 2; Length 288;  
 Best Local Similarity 30.8%; Pred. No. 3.5;  
 Matches 20; Conservative 6; Mismatches 16; Indels 23; Gaps 3;

OY 3 SRVPM-----TAMAF--HGYSYQHDLRAYGFWRLVRRFVHR-R 39  
 DB 174 SRVEMNQREKORLPRLMEDELIAMKMPVRRFRSTYTERKAHGLKRRARRVDYKTR 233

OY 40 PVYES 44  
 DB 234 RDIEA 238

RESULT 6  
 T43618  
 repA protein homolog - Yersinia pestis plasmid pCD1  
 C:Species: Yersinia pestis  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T43618  
 R:Hu, P.; Elliott, J.; McCreedy, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.; Bruba  
 J. Bacteriol. 180, 5193-5202, 1998  
 A:Title: Structural organization of virulence-associated plasmids of Yersinia pestis.  
 A:Reference number: Z22578; MUID:98422474  
 A:Accession: T43618  
 A:Molecule type: translated from GB/EMBL/DBJ  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-292 <HUP>  
 A:Cross-references: EMBL:AF053946; NID:g2996222; PIDN:AAC62601.1; PID:g2996278  
 A:Experimental source: strain KIM  
 C:Genetics:  
 A:Genome: plasmid pCD1  
 A:Note: repA  
 C:Superfamily: repl protein

Query Match 23.2%; Score 59.5; DB 2; Length 292;

Best Local Similarity 30.8%; Pred. No. 3.6;  
Matches 20; Conservative 6; Mismatches 16; Indels 23; Gaps 3;

OY 3 SRPWF-----TAAAF--HGYSYQHDRLAYGFWRLVRFVIR-R 39  
Db 178 SRVEMENQREKORLPRLNDELIAKAMRFVRRFSYQTERKAKHGILKARARRDVRDTR 237

OY 40 PVFES 44  
Db 238 RDIET 242

RESULT 7  
C70905  
hypothetical protein RV0176 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: C70905  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A: Reference number: A70500; MUID: 98295987  
A: Accession: C70905  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1322 <COL>  
A: Cross-references: GB:297050; GB:AL123456; NID:g3256008; PIDN:CAB09743.1; PID:g2213513  
A: Experimental source: strain H37RV  
C: Genetics:  
A: Gene: RV0176  
C: Superfamily: Mycobacterium tuberculosis hypothetical protein RV0176

Query Match 23.0%; Score 59; DB 2; Length 322;  
Best Local Similarity 44.1%; Pred. No. 4.6;  
Matches 15; Conservative 1; Mismatches 14; Indels 4; Gaps 1;

OY 8 TAW-----AFHGYSYQHDRLAYGFWRLVRFVIR 37  
Db 83 TGWSIGRALTGIRVRRDGSALGPRLVRLDLAH 116

RESULT 8  
T20346  
pyruvate carboxylase (EC 6.4.1.1) D2023.2 [similarity] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Aug-2000  
C:Accession: T20346  
R: Kershaw, J.  
submitted to the EMBL Data Library, October 1996  
A: Reference number: Z19260  
A: Accession: T20346  
A: Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA  
A: Residues: 1-1175 <MUI>  
A: Cross-references: EMBL:Z81052; PIDN:CAB02872.1; GSPDB:GN00023; CESP:D2023.2  
A: Experimental source: clone D2023  
C: Genetics:  
A: Gene: CESP:D2023.2  
A: Map position: 5  
A: Introns: 34/3; 103/3; 246/1; 451/3; 567/2; 821/1; 1045/3; 1126/3  
C: Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binding  
C: Keywords: ligase  
F:1140/Binding site: biotin (lys) (covalent) #status predicted

Query Match 22.6%; Score 58; DB 2; Length 1175;  
Best Local Similarity 43.3%; Pred. No. 23;  
Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

OY 16 RSYQHDRLAYGFWRLVRFVIRRRPHEVSQ 45  
Db 214 RSYSEAOAARFDGSLPFVKFVERPRHIEVQ 243

RESULT 9  
A47255  
pyruvate carboxylase (EC 6.4.1.1) precursor [similarity] - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Feb-2002  
C:Accession: A47255  
R: Zhang, J.; Xia, W.L.; Brew, K.; Ahmad, F.  
Proc. Natl. Acad. Sci. U.S.A. 90, 1766-1770, 1993  
A: Title: Adipose pyruvate carboxylase: amino acid sequence and domain structure deduced  
A: Reference number: A47255; MUID: 93189578  
A: Accession: A47255  
A: Status: preliminary  
A: Molecule type: mRNA  
A: Residues: 1-1178 <ZHA>  
A: Cross-references: GB:I09192; NID:g293743; PIDN:AAA39737.1; PID:g293744  
A: Experimental source: 3T3-L1 adipocytes  
A: Note: sequence extracted from NCBI backbone (NCBIN:126874, NCBI:126875)  
C: Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-bind  
C: Keywords: biotin binding; ligase; mitochondrion  
F:1-20/Domain: transit peptide (mitochondrion) #status predicted <TRP>  
F:21-1178/Product: pyruvate carboxylase #status predicted <MAT>  
F:39-494/Domain: biotin carboxylase homology <BCH>  
F:1105-1178/Domain: lipoyl/biotin-binding homology <LPB>  
F:1144/Binding site: biotin (lys) (covalent) #status predicted

Query Match 22.6%; Score 58; DB 1; Length 1178;  
Best Local Similarity 31.7%; Pred. No. 23;  
Matches 13; Conservative 6; Mismatches 14; Indels 8; Gaps 1;

OY 13 HGYSYQHDRLAYGFWRLVRFVIRRRPHEVSQ 45  
Db 208 HSYEELDENYTRAYSEALAAFGALGFVKFIERPRHIEVQ 248

RESULT 10  
JC2460  
pyruvate carboxylase (EC 6.4.1.1) precursor - human  
N: Alternate names: pyruvate:carbon dioxido ligase (ADP-forming)  
C:Species: Homo sapiens (man)  
C>Date: 10-Sep-1999 #sequence\_revision 18-Aug-2000 #text\_change 01-Feb-2002  
C:Accession: G01933; JC2460; B27883; S01469  
R: Walker, M.E.; Jitrapakdee, S.; Val, D.L.; Wallace, J.C.  
submitted to the EMBL Data Library, July 1995  
A: Reference number: H00708  
A: Accession: G01933  
A: Status: translated from GB/EMBL/DBJ  
A: Molecule type: mRNA  
A: Residues: 1-1178 <MAL>  
A: Cross-references: EMBL:U30891; NID:g1101028; PIDN:AAA82937.1; PID:g1101029  
R: Mackay, N.; Rigat, B.; Douglas, C.; Chen, H.S.; Robinson, B.H.  
Biochem. Biophys. Res. Commun. 202, 1009-1014, 1994  
A: Title: cDNA cloning of human kidney pyruvate carboxylase.  
A: Reference number: JC2460; MUID: 94324922  
A: Accession: JC2460  
A: Molecule type: mRNA  
A: Residues: 1-224, 'WP', 227-351, 'A', 353-384, 'PT', 387-485, 'DV', 488-637, 'R', 639-728, 'A',  
A: Cross-references: GB:S72370; NID:g652807; PIDN:AA31500.1; PID:g652808  
R: Lamhonwah, A.M.; Qian, F.; Gravel, R.A.  
Arch. Biochem. Biophys. 254, 631-636, 1987  
A: Title: Sequence homology around the biotin-binding site of human propionyl-CoA carb  
A: Reference number: A27883; MUID: 87212051  
A: Accession: B27883  
A: Molecule type: mRNA  
A: Residues: 1083-1178 <LAM>  
A: Cross-references: GB:M26122; NID:g189657; PIDN:AAA3423.1; PID:g387003  
R: Freytag, S.O.; Collier, K.J.  
J. Biol. Chem. 259, 12831-12837, 1984

```

A>Title: Molecular cloning of a cDNA for human pyruvate carboxylase. Structural relationship
A:Reference number: S01469; MUID:85030380
A:Accession: S01469
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1135-1178 <PRE>
A:Cross-references: EMBL:K02282; NID:g189655; PIDN:AAA60033.1; PID:g189656
C:Genetics:
A:Gene: GDB:PC
A:Cross-references: GDB:119472; OMIM:266150
A:Map position: 11q11-11q13.1
C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binding
C:Keywords: biotin binding; gluconeogenesis; ligase; mitochondrion
F:1-20/Domain: transit peptide (mitochondrion) #status predicted <TMP>
F:39-494/Domain: pyruvate carboxylase #status predicted <MAT>
F:1105-1178/Domain: lipoyl/biotin-binding homology <LBP>
F:1144/Binding site: biotin (Lys) (covalent) #status predicted

Query Match      22.6%; Score 58; DB 1; Length 1178;
Best Local Similarity 31.7%; Pred. NO. 23;
Matches 13; Conservative 6; Mismatches 14; Indels 8; Gaps 1;

QY      13 HGY-----RSYQHDLRANGFWRLLVRRFYHRRPHVESQ 45
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db       208 HSYEELENYTRAYSEALAFNGALFVEKFEKPRHIEVQ 248

RESULT 11
JC4391
Pyruvate carboxylase (EC 6.4.1.1) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Jan-1996 #sequence_revision 25-Apr-1997 #text_change 11-Jan-2002
C:Accession: S68252; S72393; JC4391; S06440
R:Jitrapakdee, S.; Booker, G.W.; Cassidy, A.I.; Wallace, J.C.
Biochem. J. 316, 631-637, 1996
A>Title: Cloning, sequencing and expression of rat liver pyruvate carboxylase.
A:Reference number: S68252; MUID:96257760
A:Accession: S68252
A:Molecule type: mRNA
A:Residues: 1-1178 <JTT1>
A:Cross-references: EMBL:U36585; NID:g1040973; PIDN:AA05268.1; PID:g1040974
A:Accession: S72393
A:Molecule type: Protein
A:Residues: 489-505 <JIT2>
A:Experimental source: Liver
R:Lehm, D.A.; Moran, S.M.; MacDonald, M.J.
Gene 165, 331-332, 1995
A>Title: The sequence of the rat pyruvate carboxylase-encoding cDNA.
A:Reference number: JC4391; MUID:96096548
A:Accession: JC4391
A:Molecule type: mRNA
A:Residues: 1-221, 'P', 223-865, 'D', 867-976, 'G', 978-1178 <LEH>
A:Cross-references: GB:U02314; NID:g929987; PIDN:AAA96256.1; PID:g929988
A:Experimental source: Liver
R:Thampy, K.G.; Huang, W.Y.; Wakil, S.J.
Arch. Biochem. Biophys. 266, 270-276, 1988
A>Title: A rapid purification method for rat liver pyruvate carboxylase and amino acid s
A:Reference number: S06440; MUID:89024676
A:Accession: S06440
A:Molecule type: Protein
A:Residues: 'SG', 23-25, 'PL', 28-29, 'LL', 32-34, 'P', 1134, 'A', 1136-1137, 1139-1152, 'T', 1154-1155
C:Comment: This enzyme is located in the mitochondrial matrix and catalyzes the convers
Intermediates that exit the mitochondrion for consumption in various pathways.
C:Genetics:
A:Gene: pc
A:Genome: nuclear
C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binding
C:Keywords: biotin binding; gluconeogenesis; homotetramer; ligase; mitochondrion
F:1-20/Domain: transit peptide (mitochondrion) #status predicted <TMP>
F:21-1178/Product: pyruvate carboxylase #status predicted <MAT>
F:39-494/Domain: biotin carboxylase homology <BOH>

```

```

F:1105-1178/Domain: lipoyl/biotin-binding homology <LPR>
F:1144/Binding site: biotin (Lys) (covalent) #status experimental

Query Match          22.6%; Score 58; DB 2; Length 1178;
Best Local Similarity 31.7%; Pred. No. 23;
Matches 13; Conservative 6; Mismatches 14; Indels 8; Gaps 1;

QY 13 HG-----RSYOHDRAGFMRLVRRFRPHRPHESQ 45
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 208 HSYEELEENYTRYSEALAFGNGALFEKFEKPRHIEVQ 248

RESULT 12
JC1231
beta-crystallin A3 - chicken
N:Comtains: beta-crystallin A1
C:Species: Gallus gallus (chicken)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
C:Accession: Jc1231; A25558; Jc1232
R:McBermolt, J.B.; Peterson, C.A.; Piatigorsky, J.
Gene 117, 193-200, 1992
A:Title: Structure and lens expression of the gene encoding chicken betaA3/A1-crystal
A:Reference number: Jc1231; MUID:92347693
A:Molecule type: DNA
A:Residues: 1-215 <MCD>
A:Cross-references: GB:M88460
A:Experimental source: eye lens
R:Peterson, C.A.; Piatigorsky, J.
Gene 45, 139-147, 1986
A:Title: Preferential conservation of the globular domains of the beta A3/A1-crystal
A:Reference number: A25558; MUID:87106807
A:Accession: A25558
A:Molecule type: mRNA
A:Residues: 1-215 <PFR>
A:Cross-references: GB:M15658; NID:g211668; PIDN:AAA48724.1; PID:g211669
C:Genetics:
A:Gene: beta-A3/A1
A:Introns: 11/1; 32/3; 72/2; 119/3; 167/2
C:Superfamily: beta-crystallin
C:Keywords: duplication; eye lens
F:1-215/Product: beta-crystallin A3 #status predicted <MAT>
F:18-215/Product: beta-crystallin A1 #status predicted <MA2>

Query Match          22.4%; Score 57.5; DB 2; Length 215;
Best Local Similarity 34.3%; Pred. No. 4.7;
Matches 12; Conservative 5; Mismatches 9; Indels 9; Gaps 2;

QY 4 RVP---WTAMAFHGRTSYOHLRA-----YGFWR 29
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 162 KIRCGAWCYOYRGYQYVLEADHGGDYKKMR 196

RESULT 13
A34337
propiophyl-CoA carboxylase (EC 6.4.1.3) alpha chain precursor - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 11-Jan-2002
C:Accession: A34337; A30871
R:Brown, M.F.; Taroni, F.; Szul, E.; Rosenberg, L.E.
J Biol. Chem. 264, 12680-12685, 1989
A:Title: Sequence analysis, biosynthesis, and mitochondrial import of the alpha-subunit
A:Reference number: A34337; MUID:89308706
A:Accession: A34337
A:Molecule type: mRNA
A:Residues: 'MPYRECAIRMCRCNG',1-704 <BRI>
A:Cross-references: GB:M22631
R:Brown, M.F.; Taroni, F.; Szul, E.; Rosenberg, L.E.
Submitted to GenBank, February 1989
A:Reference number: A30871
A:Accession: A30871

```

A:Molecule type: mRNA  
A:Residues: 'MYRRERFCAIRWCRNSG',1-298,'WP',301-704 <BR2>  
A:Cross-references: GB:M22631; NID:g206049; PIDN:AAA8512.1; PID:g206050  
R:Browner, M.F.; Taroni, F.; Sztul, E.; Rosenberg, L.E.  
J. Biol. Chem. 266, 4660, 1991  
A:Reference number: A43040  
A:Contents: annotation; correction  
A:Note: the first 17 residues in the original paper were derived from bacterial DNA as a  
C:Genetics:  
A:Gene: PCCA  
C:Superfamily: propionyl-CoA carboxylase alpha chain; biotin carboxylase homology; lipoyl  
C:Keywords: biotin binding; heterododecamer; ligase; mitochondrial matrix; mitochondrion  
F:1-27/Domain: transit peptide (mitochondrion) (fragment) #status predicted <TNP>  
F:28-704/Product: propionyl-CoA carboxylase alpha chain #status predicted <MAT>  
F:41-493/Domain: biotin carboxylase homology <BCH>  
F:531-704/Domain: lipoyl/biotin-binding homology <LPB>  
F:670/Binding site: biotin (lys) (covalent) #status predicted

Query Match 22.2%; Score 57; DB 2; Length 704;  
Best Local Similarity 34.1%; Pred. No. 18;  
Matches 15; Conservative 8; Mismatches 19; Indels 2; Gaps 2;  
OY 4 RVPW-TAMAFHGY-RSYQHLRAYGFWRLLVRFVRRPHVESQ 45  
Db 206 RIPWDEETRDGFRFSOEAASFQDDRLLEKFDNPHIEIQ 249

RESULT 14  
H72410  
hypothetical protein TM0162 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72410  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316  
A:Accession: H72410  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-771 <ARN>  
A:Cross-references: GB:AE001701; GB:AE000512; NID:g4980648; PIDN:AND35255.1; PID:g498065  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0162  
C:Superfamily: Thermotoga maritima hypothetical protein TM0162

Query Match 22.2%; Score 57; DB 2; Length 771;  
Best Local Similarity 36.7%; Pred. No. 20;  
Matches 11; Conservative 6; Mismatches 13; Indels 0; Gaps 0;  
OY 15 YRSYQHLRAYGFWRLLVRFVRRPHVES 44  
Db 268 YSQYFPLAHGTGLIVRLQRRHPEVNN 297

RESULT 15  
D81950  
probable transmembrane efflux protein NMA1022 [imported] - Neisseria meningitidis (stra  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: D81950  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556  
A:Accession: D81950  
A>Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-459 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84291.1; PID:g737  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA1022  
C:Superfamily: conserved hypothetical protein H11612

Query Match 22.0%; Score 56.5; DB 2; Length 459;  
Best Local Similarity 32.5%; Pred. No. 14;  
Matches 13; Conservative 6; Mismatches 12; Indels 9; Gaps 2;  
OY 2 VSRVP-----WTAMAFHGY-RSYQHLRAYGFWRLLV 32  
Db 392 VTKVPMFTHAAAFNCCGLLPGYLLAYRDMGIGYGTWALI 431

Search completed: August 15, 2002, 12:00:14  
Job time: 204 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:59:45 ; Search time 16.31 Seconds

(Without alignments)  
106.829 Million cell updates/sec

Title: DEVI-613-COMB

Sequence: 1 TVSRVPMWTAMAFHGYSYQH.....GFWRLLVRFVHRPRHVESQ 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66.5	25.9	248	1	CRB1_RAT
2	63	24.5	257	1	ICEE_MOUSE
3	60.5	23.5	252	1	CRB1_BOVIN
4	58.5	22.8	251	1	CRB1_HUMAN
5	58	22.6	1178	1	PYC_HUMAN
6	58	22.6	1178	1	PYC_MOUSE
7	57.5	22.4	215	1	CRBA_CHICK
8	57	22.2	704	1	PCCA_RAT
9	57	22.2	1178	1	PYC_RAT
10	56.5	22.0	459	1	NORM_NEIMA
11	56	21.8	242	1	ICEE_HUMAN
12	56	21.8	278	1	YH22_YEAST
13	54	21.0	198	1	CH11_RANCA
14	54	21.0	198	1	CH12_RANCA
15	54	21.0	198	1	CRBA_RANTE
16	54	21.0	215	1	CB31_RANCA
17	54	21.0	215	1	CB32_RANCA
18	54	21.0	976	1	XPR6_YARLI
19	53	20.6	177	1	CRBA_RAT
20	53	20.6	196	1	CRBB_BOVIN
21	53	20.6	196	1	CRBB_HUMAN
22	53	20.6	196	1	CRBB_MOUSE
23	53	20.6	215	1	CRBA_BOVIN
24	53	20.6	215	1	CRBA_HUMAN
25	53	20.6	238	1	CRB1_CHICK
26	53	20.6	285	1	REP2_ECOLI
27	51.5	20.0	459	1	NORM_NEIMB
28	51	19.8	422	1	EF1G_PPRUV
29	51	19.8	648	1	BGLR_CERAE
30	51	19.8	651	1	BGLR_CANPA
31	51	19.8	651	1	BGLR_FELCA
32	51	19.8	651	1	BGLR_HUMAN
33	51	19.8	791	1	G6PE_HUMAN

34	51	19.8	990	1	NFRA_ECOLI	p31600 escherichia
35	51	19.8	1203	1	YT41_CAEEL	Q11069 caenorhabdi
36	50.5	19.6	167	1	ELBS_ADE40	P10543 human adeno
37	50.5	19.6	629	1	HAP1_RAT	P54256 rattus norv
38	50.5	19.6	703	1	PCCA_HUMAN	P05165 homo sapien
39	50	19.5	196	1	CRBB_CHICK	P55164 homo sapien
40	50	19.5	343	1	GUNC_CLOSF	P23340 clostridium
41	50	19.5	343	1	GUNC_CLOTFM	P07985 clostridium
42	50	19.5	513	1	MANB_BACSM	P16699 bacillus sp
43	49.5	19.3	3859	1	RPOA_DELV	Q04561 lelystad vi
44	49	19.1	361	1	HA1A_RABIT	P01894 oryctolagus
45	49	19.1	361	1	HA1B_RABIT	P06140 oryctolagus

## ALIGNMENTS

RESULT	ID	CRB1_RAT	STANDARD:	PRT:	248 AA.
AC	P02523	CRB1_RAT			
DT	21-JUL-1986	(Rel. 01, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Beta crystallin B1 [Contains: Beta B1b crystallin].				
GN	CRYBB1.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=85175137; PubMed=3879970;				
RA	den Dunnen J.T., Moormann R.J.M., Schoenmakers J.G.G.;				
RT	"Intion insertions and deletions in the beta/gamma-crystallin gene				
RT	family: the rat beta B1 gene."				
RL	Proc. Natl. Acad. Sci. U.S.A. 83:2855-2859(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=86206054; PubMed=3458246;				
RA	den Dunnen J.T., Moormann R.J.M., Lubsen N.H., Schoenmakers J.G.G.;				
RT	"Intion insertions and deletions in the beta/gamma-crystallin gene				
RT	family: the rat beta B1 gene."				
RL	Proc. Natl. Acad. Sci. U.S.A. 83:2855-2859(1986).				
CC	-I- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS				
CC	OF THE VERTEBRATE EYE LENS.				
CC	-I- SUBUNIT: HOMO/HETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE				
CC	STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED				
CC	THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS.				
CC	-I- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR				
CC	VERY SIMILAR GREEK KEY MOTIFS.				
CC	-I- PRIM. SPECIFIC CLEAVAGES IN THE N-TERMINAL ARM OCCUR DURING LENS				
CC	MATURATION AND GIVE RISE TO TRUNCATED FORMS, LEADING TO IMPAIRED				
CC	OLIGOMERIZATION AND PROTEIN INSOLUBILIZATION. THE PROTEASE				
CC	RESPONSIBLE FOR THIS PARTIAL DEGRADATION COULD BE CALPAIN II.				
CC	-I- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL: X06377; CAA29679.1; -				
DR	EMBL: X05900; CAA29329.1; -				
DR	EMBL: M13534; AAA40979.1; -				
DR	EMBL: M13527; AAA40979.1; JOINED.				
DR	EMBL: M13528; AAA40979.1; JOINED.				
DR	EMBL: M13530; AAA40979.1; JOINED.				
DR	EMBL: M13532; AAA40979.1; JOINED.				
DR	PIR: A02925; CYRBT1.				

DR HSP: P02522; 1BDB.  
 DR Interpro: IPR001064; Crystallin.  
 DR Pfam: PF00030; crystall; 2.  
 DR PRINTS: PR01367; BGCYSTALLIN.  
 DR SMART: SM00247; XTALB; 2.  
 DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMA; 4.  
 KW Eye lens protein; Repeat; Acetylation.  
 FT INIT\_MET 0  
 FT CHAIN 1 248  
 FT MOD\_RES 1 248  
 FT DOMAIN 1 54  
 FT DOMAIN 55 94  
 FT DOMAIN 95 140  
 FT DOMAIN 141 144  
 FT DOMAIN 145 186  
 FT DOMAIN 187 230  
 FT DOMAIN 231 248  
 FT SEQUENCE 248 AA; 27912 MW; 215D6825D2FD895F CRC64;

Query Match 25.9%; Score 66.5; DB 1; Length 248;  
 Best Local Similarity 33.3%; Pred. No. 0.17;  
 Matches 18; Conservative 5; Mismatches 20; Indels 11; Gaps 2;

OY 1 TVSRVETAMAFHGYRSGH-----DLRAYGEWRL-----VRFVHRPHVE 43  
 DB 183 TVSSGTWVGXYPGYRGYQYLLEPGDFRHNWEGAFOPQMAVRRLRQWQHE 236

RESULT 2  
 ICEE\_MOUSE STANDARD; PRT; 257 AA.  
 AC 089094;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Caspase-14 precursor (EC 3.4.22.-) (CASP-14) (Mini-ICE) (MICE).  
 GN CASP14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=99040667; PubMed=9823333;  
 RA Ahmed M., Srinivasula S.M., Hegde R., Mukattash R.,  
 RA Fernandes-Alnemri T., Alnemri E.S.;  
 RT "Identification and characterization of murine caspase-14, a new  
 member of the caspase family.";  
 RL Cancer Res. 58:5201-5205(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Embryo;  
 RX MEDLINE=99222069; PubMed=10203698;  
 RA Van den Craen M., Van Loo G., Pype S., Van Griekinge W.,  
 RA Vandenabeele P.;  
 RT "Identification of a new caspase homologue: caspase-14.";  
 RL Cell Death Differ. 5:838-846(1998).  
 RN [3]  
 RP CHARACTERIZATION, AND MUTAGENESIS OF CYS-136.  
 RX Hu S., Snipas S.J., Vincenz C., Salvesen G., Dixit V.M.;  
 RT J. Biol. Chem. 273:29648-29653(1998).  
 CC -I- FUNCTION: SEEMS TO BE INVOLVED IN THE DEATH RECEPTOR AND GRANZYME  
 B APOPTOTIC PATHWAYS. MAY FUNCTION AS A DOWNSTREAM SIGNAL  
 TRANSDUCER OF CELL DEATH. MAY PLAY A ROLE IN ONTOGENESIS AND SKIN  
 PHYSIOLOGY.  
 CC -I- SUBUNIT: MAY DIMERIZE WITH LARGE PRODOMIN CASPASES.  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC

CC -I- TISSUE SPECIFICITY: EMBRYO, ADULT LIVER AND LESS IN ADULT BRAIN  
 CC AND KIDNEY.  
 CC -I- PTM: CLEAVAGE BY GRANZYME B, CASPASE-8 AND -10 GENERATES THE  
 CC TWO ACTIVE SUBUNITS. ADDITIONAL PROCESSING OF THE PROPEPTIDES IS  
 CC LIKELY DUE TO THE AUTOCATALYTIC ACTIVITY OF THE ACTIVATED  
 CC PROTEASE.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14, ALSO KNOWN AS THE  
 CC CASPASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF092997; AAC63364.1; -;  
 DR EMBL: AJ007750; CAA07678.1; -;  
 DR HSP: P42574; ICP3.  
 DR MGD: MGI:1335092; Casp14.  
 DR Interpro: IPR003576; Caspase.  
 DR Interpro: IPR002138; ICE\_p10.  
 DR Interpro: IPR001309; ICE\_p20.  
 DR Pfam: PF0655; ICE\_p10; 1.  
 DR Pfam: PF0656; ICE\_p20; 1.  
 DR SMART: SM00115; CASC; 1.  
 DR PROSITE: PS01121; CASPASE\_HIS; FALSE\_NEG.  
 DR PROSITE: PS01122; CASPASE\_CYS; FALSE\_NEG.  
 DR PROSITE: PS0207; CASPASE\_P10; 1.  
 DR PROSITE: PS0208; CASPASE\_P20; 1.  
 KW Hydrolyase; Thiol protease; Apoptosis; Zymogen.  
 FT PROPEP 1 7  
 FT CHAIN 8 2  
 FT CHAIN ? 257  
 FT ACT\_SITE 93 93  
 FT ACT\_SITE 136 136  
 FT MUTAGEN 136 136  
 FT SEQUENCE 257 AA; 29458 MW; A228D86DFBA0EB84 CRC64;  
 Query Match 24.5%; Score 63; DB 1; Length 257;  
 Best Local Similarity 42.3%; Pred. No. 0.49;  
 Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
 OY 14 GYRSYOHDLRAYGEWRLVRFVHR 39  
 DB 189 GYLSYRHDKSGSFIOTLIDVFIHK 214  
 RESULT 3  
 CRBL\_BOVIN STANDARD; PRT; 252 AA.  
 AC P07318;  
 DT 01-APR-1988 (Rel. 07; Created)  
 DT 01-NOV-1990 (Rel. 16; Last sequence update)  
 DT 01-NOV-1990 (Rel. 35; Last annotation update)  
 DE Beta crystallin Bl.  
 GN CRYB1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85134866; PubMed=6527379;  
 RA Quax-Jeukens Y., Janssen C., Quax W.J., van den Heuvel R.,  
 RT Bloemendaal H.;  
 RT "Bovine beta-crystallin complementary DNA clones. Alternating  
 RT proline/alanine sequence of beta B1 subunit originates from a  
 RT repetitive DNA sequence.";  
 RT

```

RL J. Mol. Biol. 180:457-472(1984).
RN [2]
RP SEQUENCE.
RC TISSUE=Iens cortex;
RX MEDLINE=84132067; Pubmed=6696025;
RA Berbers G.A.M., Hoekman W.A., Bloemendal H., de Jong W.W.,
RT Kleinschmidt T., Braunitzer G.;
RT "Homology between the primary structures of the major bovine beta-
RT crystallin chains.";
RL Eur. J. Biochem. 139:467-479(1984).
CC -1- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC OF THE VERTEBRATE EYE LENS.
CC -1- SUBUNIT: HOMO/HETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE
CC STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED
CC THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS.
CC -1- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR
CC VERY SIMILAR GREEK KEY MOTIFS.
CC -1- PTM: SPECIFIC CLEAVAGES IN THE N-TERMINAL ARM OCCUR DURING LENS
CC MATURATION AND GIVE RISE TO TRUNCATED FORMS, LEADING TO IMPAIRED
CC OLIGOMERIZATION AND PROTEIN INSOLUBILIZATION.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X01808; CAA25951.1; -.
DR PIR: A27898; A27898.
DR HSP: P02564; S07264.
DR InterPro: IPR001064; Crystallin.
DR Pfam: PF00030; Crystall_2.
DR PRINTS: PR01367; BCGCRYSTALLIN.
DR SMART: SM00247; XTALBd; 2.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 4.
KW Eye lens protein; Repeat; Acetylation.
FT INIT_MET 0
FT MOD_RES 1
FT DOMAIN 1 58 ACETYLATION (PROBABLE).
FT DOMAIN 59 98 N-TERMINAL ARM.
FT DOMAIN 99 144 MOTIF 1.
FT DOMAIN 145 148 MOTIF 2.
FT DOMAIN 149 190 MOTIF 3.
FT DOMAIN 191 234 MOTIF 4.
FT DOMAIN 235 252 C-TERMINAL ARM.
FT CONFLICT 7 10 ASAT -> TSAA (IN REF. 2).
FT CONFLICT 50 50 MISSING (IN REF. 2).
FT CONFLICT 93 101 MISSING (IN REF. 2).
FT CONFLICT 146 146 A -> S (IN REF. 2).
SQ SEQUENCE 252 AA; 28012 MW; 7E508660EE477D20 CRC64;

```

```

Query Match 23.5%; Score 60.5; DB 1; Length 252;
Best Local Similarity 32.1%; Pred. No. 1;
Matches 17; Conservative 5; Mismatches 20; Indels 11; Gaps 2;

```

```

QY 2 VSRVPTWMAFHGYRSTOH-----DLRAYGFWRL-----VRRFVHRPPHVE 43
Db 188 VSSSTWVGYYGPGYRGYQYLLEPDGFRHMNMKMAFQPMQMAVRLRDRQWIRE 240

```

```

GN CRYBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=66224006; Pubmed=8626774;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21 AND 24-251.
RC TISSUE=Lens;
RX MEDLINE=20057165; Pubmed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
RA Clump M., Smink L.J., Alnscoough R., Almeida J.P., Babbage A.,
RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA Blakey S., Bridgman A.M., Buck D., Burgess J., Burrill W.D.,
RA Burton V., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA Grafham D., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA Kimberley A., King A., Laird G.K., Langford C.F., Laversha M.A.,
RA Lloyd C., Lloyd D.M., Matryn I.D., Mashreghi-Mohammadi M.,
RA Matthews L., Mccann O.T., Mcclay J., McLaren S., Mcmurtry A.V.,
RA Malne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.A.,
RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA Ramsey Y., Rogers L., Rose M.T., Scott C.E., Sena H.K., Skuce C.D.,
RA Smalley S., Smith M.L., Soderlund C., Spurgeon L., Steward A.A.,
RA Whitely M.N., Willey D., Williams M., Walli J.M.,
RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA Beck S., Rogers J., Shimizu N., Minooshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kudo J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA Lai H., Lao H.I., Lewis S., Lin S.-P., Loh P., Mala J.E.,
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozerky P.,
RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K.,
RA Nelson J., Korf I., Bedell J.A., Hillier L., Maris E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kutahshi H., Salta S.,
RA Budar M.L., Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Serousi E., Fransson I., Tapla I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tiliahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [3]
RP SEQUENCE OF 199-245 FROM N.A.
RX MEDLINE=96121383; Pubmed=8575764;
RA Hulsebos T.J.M., Gilbert D.J., Delattre O., Smink L.J., Dunham I.,
RA Westerveld A., Thomas G., Jenkins N.A., Copeland N.G.;
RT "Assignment of the beta B1 crystallin gene (CRYBB1) to human
RT chromosome 22 and mouse chromosome 5.";
RL Genomics 29:712-718(1995).
CC -1- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC OF THE VERTEBRATE EYE LENS.
CC -1- SUBUNIT: HOMO/HETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE
CC STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED
CC THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS.
CC -1- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR

```

```

CC      VERY SIMILAR GREEK KEY MOTIFS.
CC      -1- PTM: SPECIFIC CLEAVAGES IN THE N-TERMINAL ARM OCCUR DURING LENS
CC      MATURATION AND GIVE RISE TO TRUNCATED FORMS, LEADING TO IMPAIRED
CC      OLIGOMERIZATION AND PROTEIN INSOLUBILIZATION.
CC      -1- MASS SPECTROMETRY: MW=27941; MW_ERR=6; METHOD=Electrospray.
CC      -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRISTALLIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U35340; AAC50383.1; -
DR      EMBL: 295115; CAB08268.1; -
DR      EMBL: X86398; CAA60150.1; -
DR      HSSP: P02522; 1BLB.
DR      MIM: 600929; -.
DR      InterPro: IPR001064; Crystallin.
DR      Pfam: PF00030; Crystallin.
DR      PRINTS: PR01367; BGCYSTALLIN.
DR      SMART: SM00247; XTALBq; 2.
DR      PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 4.
KM      Eye lens protein; Repeat: Acetylation.
FT      INIT_MET 0
FT      MOD_RES 1 1 ACETYLATION.
FT      DOMAIN 1 57 N-TERMINAL ARM.
FT      DOMAIN 58 97 MOTIF 1.
FT      DOMAIN 98 143 MOTIF 2.
FT      DOMAIN 144 147 CONNECTING PEPTIDE.
FT      DOMAIN 148 189 MOTIF 3.
FT      DOMAIN 190 233 MOTIF 4.
FT      DOMAIN 234 251 C-TERMINAL ARM.
SQ      SEQUENCE 251 AA; 27892 MW; 0AD81A8EACAA86F54 CRC64;
OY      2 VSRVPTAAAFHGRYSYOH-----DIRAYGFMTL-----VRFVHRRPHVE 43
Db      187 VSSGTVWGIGYRGRGYTLLEPGDERHNMWNGAFOPOMOSLRRLDKOMHLE 239
Query Match 22.8%; Score 58.5; DB 1; Length 251;
Best Local Similarity 28.3%; Pred. No. 1.8;
Matches 15; Conservative 8; Mismatches 19; Indels 11; Gaps 2;
RESULT 5
PYC_HUMAN ID PYC_HUMAN STANDARD: PRT; 1178 AA.
AC P1198; Q16705;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Pyruvate carboxylase, mitochondrial precursor (EC 6.4.1.1) (Pyruvic
DE carboxylase) (PCB).
GN PC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver, and Kidney;
RX MEDLINE=95002202; PubMed=7918683;
RA Wexler I.D., Du Y., Lisgaris M.V., Mandal S.K., Freytag S.O.,
RA Yang B.-S., Liu T.-C., Kwon M., Patel M.S., Kerr D.S.;
RT "Primary amino acid sequence and structure of human pyruvate
RT carboxylase."
RL Biochim. Biophys. Acta 1227:46-52(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;

```

```

RX MEDLINE=94324922; PubMed=8048912;
RA Mackay N., Rigat B., Douglas C., Chen H.S., Robinson B.H.;
RT "cDNA cloning of human kidney pyruvate carboxylase."
RL Biochem. Biophys. Res. Commun. 202:1009-1014(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver, and Kidney;
RA Walker M.E., Jitrapakdee S., Val D.L., Wallace J.C.;
RL Submitted (Jul-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1083-1178 FROM N.A.
RX MEDLINE=87212051; PubMed=3555348;
RA Lamtonwan A.-W., Quan F., Gravel R.A.;
RT "Sequence homology around the biotin-binding site of human
RT propionyl-CoA carboxylase and pyruvate carboxylase."
RL Arch. Biochem. Biophys. 254:631-636(1987).
RN [5]
RP SEQUENCE OF 1135-1178 FROM N.A.
RX MEDLINE=85303380; PubMed=6548474;
RA Freytag S.O., Collier K.J.;
RT "Molecular cloning of a cDNA for human pyruvate carboxylase.
RT Structural relationship to other biotin-containing carboxylases and
RT regulation of mRNA content in differentiating preadipocytes."
RL J. Biol. Chem. 259:12831-12837(1984).
RN [6]
RP VARIANTS PC DEFICIENCY THR-610 AND ILE-743.
RX MEDLINE=98254451; PubMed=9585612;
RA Carbone M.A., Mackay N., Ling M., Cole D.E.C., Douglas C., Rigat B.,
RA Feigenbaum A., Clarke J.T.R., Haworth J.C., Greenberg C.R.,
RA Seargent L., Robinson B.H.;
RT "Amerindian pyruvate carboxylase deficiency is associated with two
RT distinct missense mutations."
RL Am. J. Hum. Genet. 62:1312-1319(1998).
CC -1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
CC INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
CC ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CC CARBOXYL GROUP TO PYRUVATE IN THE SECOND. CATALYZES IN A TISSUE
CC SPECIFIC MANNER, THE INITIAL REACTIONS OF GLUCOSE (LIVER, KIDNEY)
CC AND LIPID (ADIPOSE TISSUE, LIVER, BRAIN) SYNTHESIS FROM PYRUVATE.
CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + Hco(3)(-) = ADP + phosphate +
CC oxaloacetate.
CC -1- COFACTOR: BIOTIN AND MANGANESE.
CC -1- PATHWAY: GLUCONEOGENESIS AND LIPOGENESIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- DISEASE: DEFICIENCY IN PC CAUSES LACTIC ACIDOSIS, MENTAL
CC RETARDATION AND DEATH. OCCURS IN THREE FORMS: TYPE A (MILD); TYPE
CC B (SEVERE NEONATAL) AND A VERY MILD LACTICACIDEMIA.
CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES
CC AND CARBAMYL PHOSPHATE SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR      EMBL: U04641; AAA99537.1; -
DR      EMBL: 572370; AAB31500.1; -
DR      EMBL: U30891; AAA82937.1; -
DR      EMBL: M26122; AAA36423.1; -
DR      EMBL: K02282; AAA60033.1; -
DR      PIR: B27883; B27883.
DR      PIR: S01469; S01469.
DR      HSSP: P24182; 1BNC.
DR      MIM: 266150; -.
DR      InterPro: IPR001882; Biotin.
DR      InterPro: IPR000089; Biotin_1lpoyl.
DR      InterPro: IPR000901; CPSase.
DR      InterPro: IPR000891; HMGL-like.
DR      InterPro: IPR003379; PYC_OMDA.

```

RESULT	6
CC	PCYC_MOUSE
ID	PTC_MOUSE
AC	STANDARD;
CC	005920;
DT	01-FEB-1994 (Rel. 28, Created)
DT	01-FEB-1994 (Rel. 28, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Pyruvate carboxylase, mitochondrial precursor (EC 6.4.1.1) (Pyruvic carboxylase) (PCB).
DE	PC OR PCX.
OS	Mus musculus (Mouse).
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=adipocyte;
RX	MEDLINE=93189578; Pubmed=8446588;
RA	zheng J., Xia W.L., Brew K., Ahmad F.;
RT	"Adipose pyruvate carboxylase: amino acid sequence and domain structure deduced from cDNA sequencing."
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1766-1770(1993).
CC	-1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
CC	INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
CC	ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CC	CARBOXYL GROUP TO PYRUVATE IN THE SECOND. CATALYZES IN A TISSUE
CC	SPECIFIC MANNER, THE INITIAL REACTIONS OF GLUCOSE (LIVER, KIDNEY)
CC	AND LIPID (ADIPOSE TISSUE, LIVER, BRAIN) SYNTHESIS FROM PYRUVATE.
CC	-1- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3)(-) = ADP + phosphate +
CC	oxaloacetate.
CC	-1- COFACTOR: BIOTIN AND MANGANESE.

RESULT	7			
CRBA_CHICK		STANDARD;	PRT;	215 MA.
ID	CRBA_CHICK			
AC	P10042;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Beta crystallin A3 [Contains: Beta crystallin A1].			
GN	CRYBA1.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Atherosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae			
OC	Gallus.			
DX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87106807; Pubmed=3803920;			

```

RA Peterson C.A., Plattigorsky J.;
RT "Preferential conservation of the globular domains of the beta A3/A1-
RL crystallin polypeptide of the chicken eye lens.";
RN Gene 45:139-147(1986).
RP SEQUENCE FROM N.A.
RX MEDLINE-92347693; PubMed1353472;
RA McDermott J.B., Peterson C.A., Plattigorsky J.;
RT "Structure and lens expression of the gene encoding chicken beta
RL A3/A1-crystallin.";
RN Gene 117:193-200(1992).
CC -!- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC OF THE VERTEBRATE EYE LENS.
CC -!- SUBUNIT: HOMO/ETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE
CC STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED
CC THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS (BY SIMILARITY).
CC -!- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR
CC VERY SIMILAR GREEK KEY MOTIFS.
CC -!- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M15658; AAA48724.1; -.
DR EMBL: M15658; AAA48725.1; ALT_INIT.
DR EMBL: M84460; AAA48612.1; ALT_INIT.
DR EMBL: M84460; AAA48611.1; ALT_SEQ.
DR PIR: A25358; A25358.
DR PIR: JCI231; JCI231.
DR HSSP: P02522; 1BLB.
DR InterPro: IPR001064; Crystallin.
DR Pfam: PF00030; crystall. 2.
DR PRINTS: PR01367; BGCCRYSTALLIN.
DR SMART: SM00247; XTALDg; 2.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 4.
KW Eye lens protein; Repeat; Alternative initiation.
FT CHAIN 1 215 BETA CRYSTALLIN A3.
FT TRANSIT 1 215 FOR BETA CRYSTALLIN A1.
FT INIT_MET 18 18 FOR BETA CRYSTALLIN A1.
FT DOMAIN 1 30 N-TERMINAL ARM.
FT DOMAIN 31 70 MOTIF 1.
FT DOMAIN 71 118 MOTIF 2.
FT DOMAIN 119 123 CONNECTING PEPTIDE.
FT DOMAIN 124 165 MOTIF 3.
FT DOMAIN 166 215 MOTIF 4.
SQ SEQUENCE 215 AA; 24749 MW; 6951BAE6826B2921 CRC64;

Query Match 22.4%; Score 57.5; DB 1; Length 215;
Best Local Similarity 34.3%; Pred. No. 2.1;
Matches 12; Conservative 5; Mismatches 9; Indels 9; Gaps 2;

OY 4 RVP--WTAMAFHGYSYOHDLRA-----YGFWR 29
DB 162 KIPGAMVCYGYRGYGVLEADHNGDYKHM 196

RESULT 8
PCCA_RAT ID STANDARD; PRT; 704 AA.
AC P14882;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Propionyl-CoA carboxylase alpha chain, mitochondrial precursor
DE (EC 6.4.1.3) (PCCAase alpha subunit) (Propanoyl-CoA:carbon dioxide
DE ligase alpha subunit) (Fragment).
GN PCCA.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-89308706; PubMed2745462;
RA Browner M.F., Taroni F., Szul E., Rosenberg L.E.;
RT "Sequence analysis, biogenesis, and mitochondrial import of the
RL alpha-subunit of rat liver propionyl-CoA carboxylase.";
RN J. Biol. Chem. 264:12680-12685(1989).
RN [2]
RP REVISIONS.
RA Browner M.F., Taroni F., Szul E., Rosenberg L.E.;
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + propanoyl-CoA + HCO(3)(-) = ADP +
CC phosphate + (S)-methylmalonyl-CoA.
CC -!- COFACTOR: BIOTIN.
CC -!- PATHWAY: KEY ENZYME IN THE CATABOLIC PATHWAY OF ODD-CHAIN
CC FATTY ACIDS, ISOLEUCINE, THREONINE, METHIONINE, AND VALINE.
CC -!- SUBUNIT: PROBABLY A DODECAMER COMPOSED OF SIX BIOTIN-CONTAINING
CC ALPHA SUBUNITS AND SIX BETA SUBUNITS.
CC -!- DISBASE: PROPRIONIC ACIDEMIA DUE TO RECESSIVELY INHERITED
CC DEFICIENCY OF PCCAASE ACTIVITY OFTEN CAUSES LIFE-THREATENING
CC KETOSIS AND ACIDOSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M22631; AAA88512.1; ALT_SEQ.
DR PIR: A34337; A34337.
DR HSSP: P24182; 1DVL.
DR InterPro: IPR001882; Biotin.
DR InterPro: IPR000089; Biotin_1lpoyl.
DR InterPro: IPR000901; CPSase.
DR Pfam: PF02785; Biotin_carb_C; 1.
DR Pfam: PF00364; biotin_1lpoyl; 1.
DR Pfam: PF00289; CPSase_L_chain; 1.
DR Pfam: PF02786; CPSase_L_D2; 1.
DR PROSITE: PS00188; BIOTIN; 1.
DR PROSITE: PS00866; CPSASE_1; 1.
DR PROSITE: PS00867; CPSASE_2; 1.
KW Mitochondrion; Ligase; Biotin; ATP-binding; Transit peptide.
FT NON_TER 1 1
FT TRANSIT 1 21
FT CHAIN 22 704 PROPRIONYL-CoA CARBOXYLASE ALPHA CHAIN.
FT NP_BIND 199 204 ATP (POTENTIAL).
FT ACT_SITE 329 329 BY SIMILARITY.
FT BINDING 670 670 BIOTIN (BY SIMILARITY).
SQ SEQUENCE 704 AA; 77711 MW; 36CEEC52DE2D2A8A CRC64;

Query Match 22.2%; Score 57; DB 1; Length 704;
Best Local Similarity 34.1%; Pred. No. 7.5;
Matches 15; Conservative 8; Mismatches 19; Indels 2; Gaps 2;

OY 4 RVPW-TAMAFHGYSYOHDLRAYGFWRLLVRRFRRHPPVESQ 45
DB 206 RLPWDEETRDGFRFSQEAASFQDRLLEKFIIDNPRIHQ 249

RESULT 9
PVC_RAT ID STANDARD; PRT; 1178 AA.
AC P52873; Q64555;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
GN PVC.

```

DE 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Pyruvate carboxylase, mitochondrial precursor (EC 6.4.1.1) (Pyruvic  
DE carboxylase) (PCB).  
GN PC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=96096548; PubMed=8522203;  
RA Lehn D.A., Moran S.M., MacDonald M.J.;  
RT "The sequence of the rat pyruvate carboxylase-encoding cDNA.";  
RL Gene 165:331-332(1995).  
RL [2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX STRAIN=Wistar; TISSUE=Liver;  
RX MEDLINE=96257760; PubMed=8687410;  
RA Jitrapakdee S., Booker G.W., Cassidy A.I., Wallace J.C.;  
RT "Cloning, sequencing and expression of rat liver pyruvate  
RT carboxylase.";  
RL Biochem. J. 316:631-637(1996).  
CC -1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,  
CC INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY  
CC ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE  
CC CARBOXYL GROUP TO PYRUVATE IN THE SECOND. CATALYZES IN A TISSUE  
CC SPECIFIC MANNER, THE INITIAL REACTIONS OF GLUCOSE (LIVER, KIDNEY)  
CC AND LIPID (ADIPOSE TISSUE, LIVER, BRAIN) SYNTHESIS FROM PYRUVATE.  
CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + Hco(3)(-) = ADP + phosphate +  
CC oxaloacetate.  
CC -1- COFACTOR: BIOTIN AND MANGANESE (BY SIMILARITY).  
CC -1- PATHWAY: GLUCONEOGENESIS AND LIPIDGENESIS.  
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES  
CC AND CARBAMYL PHOSPHATE SYNTHETASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: U32314; AAA96256.1; -;  
DR EMBL: U36585; AAC52668.1; -;  
DR HSSP: P24182; 1BNC.  
DR InterPro: IPR001882; Biotin.  
DR InterPro: IPR000089; Biotin\_1lpoyl.  
DR InterPro: IPR000901; CPSase.  
DR InterPro: IPR000891; HMGL-1like.  
DR InterPro: IPR003379; pvc OADA.  
DR Pfam: PF02785; Biotin\_carb\_C; 1.  
DR Pfam: PF00364; biotin\_1lpoyl; 1.  
DR Pfam: PF00289; CPSase\_L; 1.  
DR Pfam: PF02786; CPSase\_L\_D2; 1.  
DR Pfam: PF00682; HMGL-1like; 1.  
DR Pfam: PF02436; pvc OADA; 1.  
DR PROSITE: PS00188; BIOTIN; 1.  
KW Lipase; Multifunctional enzyme; Biotin; Manganese; Gluconeogenesis;  
KW ATP-binding; Mitochondrial enzyme; Lipid synthesis; Transil peptide.  
FT TRANSIT 1 20  
FT CHAIN 21 1178  
FT DOMAIN 550 1000  
FT DOMAIN 1096 1178  
FT  
FT NP\_BIND 198 203  
FT ACT\_SITE 328 328  
FT BINDING 1144 1144  
FT CONFLICT 222 222

FT	CONFLICT	866	866	D -> I (IN REF. 2).
FT	CONFLICT	977	977	G -> R (IN REF. 2).
SO	SEQUENCE	1178 AA;	129689 MW;	8E5FA19BC132A8DD CRC64;

Query Match Best Local Similarity 22.2%; Score 57; DB 1; Length 1178;  
Matches 13; Conservative 6; Mismatches 14; Indels 8; Gaps 1;

```

OY      13 HGY-----RSYQHDLRAYGFWMLVRRYHRRHWESQ 45
          | | | | | | | | | | | | | | | | | | | | |
DB       208 HSYEELEENYFRAYPEALAFNGALFEVEKFIETKRHIIEVQ 248

```

ID	NORM_NEIMA	STANDARD:	PRT:	459 AA.
AC	O9JCV27:			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Probable multidrug resistance protein norm (Na(+)/drug antiporter)			
DE	(Multidrug-efflux transporter).			
GN	NORM OR NMA1022.			
OS	Neisseria meningitidis (serogroup A).			
OC	Bacteria: Proteobacteriia; beta subdivision; Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=65699;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;			
RX	MEDLINE=20222556; PubMed=10761919;			
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,			
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,			
RA	Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,			
RA	Jagels K., Leather S., Moule S., Mungrall K., Quail M.A.,			
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,			
RA	Whitehead S., Spratt B.G., Barrrell B.G.;			
RT	"Complete DNA sequence of a serogroup A strain of Neisseria			
RT	meningitidis Z2491.";			
RL	Nature 404:502-506(2000).			
CC	-I- FUNCTION: FUNCTIONS AS A NA(+)/DRUG ANTIPORTER (BY SIMILARITY).			
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane			
CC	(potential).			
CC	-I- SIMILARITY: BELONGS TO THE MULTI ANTIMICROBIAL EXTRUSION (MATE)			
CC	FAMILY. NORM (FC 2. A 66.1.1) SUBFAMILY			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AL162754; CAB84291.1; "			
DR	InterPro; IPRO02528; UPF0013.			
DR	Pfam; PF01554; UPF0013; 2.			
KW	Transport; Sodium transport; Transmembrane; Inner membrane;			
KW	Complete proteome.			
FT	TRANSMEM	20	40	POTENTIAL.
FT	TRANSMEM	53	73	POTENTIAL.
FT	TRANSMEM	100	120	POTENTIAL.
FT	TRANSMEM	132	152	POTENTIAL.
FT	TRANSMEM	168	188	POTENTIAL.
FT	TRANSMEM	202	222	POTENTIAL.
FT	TRANSMEM	252	272	POTENTIAL.
FT	TRANSMEM	285	305	POTENTIAL.
FT	TRANSMEM	325	345	POTENTIAL.
FT	TRANSMEM	358	378	POTENTIAL.
FT	TRANSMEM	395	415	POTENTIAL.
FT	TRANSMEM	423	443	POTENTIAL.
SO	SEQUENCE	459 AA;	49941 MW;	B2EFDL15C8EA94E CRC64;







```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Beta crystallin Al-1.
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
NCBI_TaxID=8400;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=96192032; PubMed=8619837;
RA Lu S.-F., Pan F.-M., Chiu S.-H.;
RT "Sequence analysis of four acidic beta-crystallin subunits of
RT amphibian lenses: phylogenetic comparison between beta- and
RT gamma-crystallins."
RT Biochem. Biophys. Res. Commun. 221:219-228(1996).
RL
CC -1- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC OF THE VERTEBRATE EYE LENS.
CC -1- SUBUNIT: HOMO/HETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE
CC STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED
CC THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS (BY SIMILARITY).
CC -1- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR
CC VERY SIMILAR GREEK KEY MOTIFS.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X87759; CA61033.1; -.
DR HSSP; P02522; IBLB.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF00030; crystal1; 2.
DR PRINTS; PR01367; BGCRCRYSTALLIN.
DR SMART; SM00247; XTALBd; 2.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 4.
KW Eye lens protein; Repeat.
FT DOMAIN 1 13 N-TERMINAL ARM.
FT DOMAIN 14 53 MOTIF 1.
FT DOMAIN 54 101 MOTIF 2.
FT DOMAIN 102 106 CONNECTING PEPTIDE.
FT DOMAIN 107 148 MOTIF 3.
FT DOMAIN 149 198 MOTIF 4.
SQ SEQUENCE 198 AA; 23236 MW; 327BAFD1461F20E7 CRC64;

Query Match 21.0%; Score 54; DB 1; Length 198;
Best Local Similarity 34.5%; Pred. No. 5.3;
Matches 10; Conservative 4; Mismatches 9; Indels 6; Gaps 1;

OY 7 WTAAAFHGYSYQHDLRA-----YGFWR 29
Db 151 WCYQYPRGYRYTILSDHHGGEYKHWR 179

RESULT 14
CB12_RANCA STANDARD; PRT; 198 AA.
ID CB12_RANCA
AC Q91314;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Beta crystallin Al-2.
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
NCBI_TaxID=8400;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=96192032; PubMed=8619837;
RA Lu S.-F., Pan F.-M., Chiu S.-H.;
RT "Sequence analysis of four acidic beta-crystallin subunits of
RT amphibian lenses: phylogenetic comparison between beta- and
RT gamma-crystallins."
RT Biochem. Biophys. Res. Commun. 221:219-228(1996).
RL
CC -1- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC OF THE VERTEBRATE EYE LENS.
CC -1- SUBUNIT: HOMO/HETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE
CC STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED
CC THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS (BY SIMILARITY).
CC -1- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR
CC VERY SIMILAR GREEK KEY MOTIFS.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X87760; CA61034.1; -.
DR HSSP; P02522; IBLB.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF00030; crystal1; 2.
DR PRINTS; PR01367; BGCRCRYSTALLIN.
DR SMART; SM00247; XTALBd; 2.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 4.
KW Eye lens protein; Repeat.
FT DOMAIN 1 13 N-TERMINAL ARM.
FT DOMAIN 14 53 MOTIF 1.
FT DOMAIN 54 101 MOTIF 2.
FT DOMAIN 102 106 CONNECTING PEPTIDE.
FT DOMAIN 107 148 MOTIF 3.
FT DOMAIN 149 198 MOTIF 4.
SQ SEQUENCE 198 AA; 23096 MW; 52A09C6422C0C775 CRC64;

Query Match 21.0%; Score 54; DB 1; Length 198;
Best Local Similarity 34.5%; Pred. No. 5.3;
Matches 10; Conservative 4; Mismatches 9; Indels 6; Gaps 1;

OY 7 WTAAAFHGYSYQHDLRA-----YGFWR 29
Db 151 WCYQYPRGYRYTILSDHHGGEYKHWR 179

RESULT 15
ID CB1A_RANCA STANDARD; PRT; 198 AA.
ID CB1A_RANCA
AC P07317;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Beta crystallin Al.
OS Rana temporaria (European common frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
NCBI_TaxID=8407;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=88050966; PubMed=3499937;
RA Luchin S.V., Zhnovleva R.D., Tomarev S.I., Dolgilevich S.M.,
RA Gause G.G., Jr., Bax J.B., Driessen H.P.C., Blundell T.L.;
RT "Frog lens beta Al-crystallin: the nucleotide sequence of the cloned
RT cDNA and computer graphics modelling of the three-dimensional

```

```

RT structure.";
RL Biochim. Biophys. Acta 916:163-171(1987).
RM [2]
RP SEQUENCE FROM N.A.
RA Luchin S.V., Tomarev S.I., Dolgilevich S.M., Kraev A.S.,
RA Stryabin K.G., Gause G.G. Jr.;
RT "Isolation and structure of a clone of recombinant cDNA encoding
RT beta-crystallin of the eye lens of the frog Rana temporaria.";
RL Dokl. Biochem. 279:357-361(1985).
CC -1- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC OF THE VERTEBRATE EYE LENS.
CC -1- SUBUNIT: HOMO/HETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE
CC STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED
CC THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS (BY SIMILARITY).
CC -1- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR
CC VERY SIMILAR GREEK KEY MOTIFS.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02767; CAA26544.1; -
DR EMBL; X06421; CAA29730.1; -
DR PIR; S01608; S01608.
DR HSSP; P02522; 1BLB.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF00030; Crystall; 2.
DR PRINTS; PR01367; BGCRYSTALLIN.
DR SMART; SM00247; XTALP9; 2.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 4.
KW Eye lens protein; Repeat.
FT DOMAIN 1 13 N-TERMINAL ARM.
FT DOMAIN 14 53 MOTIF 1.
FT DOMAIN 54 101 MOTIF 2.
FT DOMAIN 102 106 CONNECTING PEPTIDE.
FT DOMAIN 107 148 MOTIF 3.
FT DOMAIN 149 198 MOTIF 4.
FT VARIANT 144 144 M -> R.
SQ SEQUENCE 198 AA; 23246 MW; CA049CDA7D6AB999 CRC64;

```

```

Query Match 21.0%; Score 54; DB 1; Length 198;
Best Local Similarity 34.5%; Pred. No. 5.3;
Matches 10; Conservative 4; Mismatches 9; Indels 6; Gaps 1;
OY 7 WTAAAFHGTRSYQHDLR-----YGFWR 29
DB 151 WVCYQYPGYRGYQYILSDHGGGEYKHM 179

```

Search completed: August 15, 2002, 12:05:55  
Job time: 370 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:59:20 ; Search time 41.42 Seconds  
(without alignments)  
187.947 Million cell updates/sec

Title: DEVI-613-COMB  
Perfect score: 257  
Sequence: 1 TVSRVPTMAAFHGYRHYQH.....GFNRLVRFVHRPRHVESQ 45

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_virus:\*  
16: sp.\_bacteriophage:\*  
17: sp.\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66.5	25.9	245	11 Q9JHV9	Q9JHV9 ratu...
2	65.5	25.5	250	11 Q9WVJ5	Q9WVJ5 mus...
3	61.5	23.9	537	10 Q9SD75	Q9SD75 arabidopsi...
4	61	23.7	724	5 Q19842	Q19842 caenorhabdi...
5	59.5	23.2	288	2 Q52214	Q52214 yersinia en...
6	59.5	23.2	288	2 Q9R3G9	Q9R3G9 yersinia pe...
7	59.5	23.2	292	2 Q68716	Q68716 yersinia pe...
8	59	23.0	322	16 Q07420	Q07420 mycobacteri...
9	58.5	22.8	336	4 Q9H495	Q9H495 homo sapien...
10	58.5	22.8	397	4 Q9H497	Q9H497 homo sapien...
11	58.5	22.8	397	4 Q9H6E7	Q9H6E7 homo sapien...
12	58	22.6	1175	5 Q17732	Q17732 caenorhabdi...
13	57	22.2	290	2 Q52218	Q52218 klebsiella...
14	57	22.2	771	16 Q9WY09	Q9WY09 thermotoga...
15	57	22.2	1180	13 Q9DDT1	Q9DDT1 brachydanio...
16	56.5	22.0	77	4 Q9BX23	Q9BX23 homo sapien...

17	56	21.8	285	2	Q9S121	Q9S121 escherichia	
18	55.5	21.6	232	13	Q90WT1	Q90WT1 brachydanio	
19	55.5	21.6	300	3	Q94457	Q94457 schizosacch	
20	55	21.4	285	2	Q9AJU6	Q9AJU6 shigella fl	
21	55	21.4	285	2	Q47618	Q47618 escherichia	
22	55	21.4	289	2	Q9AFJ7	Q9AFJ7 shigella fl	
23	54.5	21.2	271	2	Q9FCG3	Q9FCG3 streptomyce	
24	54.5	21.2	1929	5	Q95Y09	Q95Y09 leishmania	
25	54	21.0	339	2	Q9KWB6	Q9KWB6 agrobacteri	
26	54	21.0	398	16	Q9PQ84	Q9PQ84 ureaplasma	
27	54	21.0	441	16	Q9Z0A4	Q9Z0A4 rhizobium m	
28	54	21.0	631	5	Q9VEG1	Q9VEG1 drosophila	
29	54	21.0	1583	10	Q9S7A7	Q9S7A7 oryza sativ	
30	53.5	20.8	227	13	Q9W6T4	Q9W6T4 xenopus lae	
31	53.5	20.8	287	2	Q9ZBW5	Q9ZBW5 klebsiella	
32	53.5	20.8	291	10	Q80435	Q80435 arabidopsi	
33	53.5	20.8	1359	4	Q96U65	Q96U65 homo sapien	
34	53	20.6	32	6	Q28257	Q28257 canis fami	
35	53	20.6	108	10	Q9AX84	Q9AX84 oryza sativ	
36	53	20.6	198	6	Q9SKK5	Q9SKK5 oryctolagus	
37	53	20.6	37	198	6	Q9SKK5	Q9SKK5 oryctolagus
38	53	20.6	215	6	Q9SKK6	Q9SKK6 oryctolagus	
39	53	20.6	215	11	Q91XW4	Q91XW4 cavia porce	
40	53	20.6	285	2	Q82926	Q82926 escherichia	
41	53	20.6	285	2	Q930L1	Q930L1 escherichia	
42	53	20.6	326	2	Q9KZK7	Q9KZK7 streptomyce	
43	53	20.6	705	10	Q9SU64	Q9SU64 arabidopsi	
44	53	20.6	966	16	Q9ZKQ3	Q9ZKQ3 rhizobium m	
45	52.5	20.4	211	10	Q94419	Q94419 arabidopsi	

## ALIGNMENTS

RESULT 1  
ID Q9JHV9 PRELIMINARY; PRT; 245 AA.  
AC Q9JHV9:  
DT 01-OCT-2000 (TREMREL. 15, Created)  
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)  
DE BETAB1-CRYSTALLIN (FRAGMENT).  
OS CRYBB1.  
GN Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-SPRAGUE-DAWLEY; TISSUE=LENS;  
RA Lampi K.J., Shih M., Shearer T.R., David L.L.;  
RT "Lens proteomics: Analysis of rat crystallin sequences and two-dimensional electrophoresis map."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS OF THE VERTEBRATE EYE LENS (BY SIMILARITY).  
CC -!- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR VERY SIMILAR GREEK KEY MOTIFS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.  
CC EMBL: AF286652; AAF97950.1; -.  
DR HSSP: P02522; 2BB2.  
DR InterPro: IPR001064; Crystallin.  
DR PRINTS: PR01367; BGCRCRYSTALLIN.  
DR SMART: SM00247; XTBAB7.2.  
DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; 4.  
KW duplication; Eye lens protein.  
FT NON\_TER 1  
SQ SEQUENCE 245 AA; 27576 MW; E1565A564FCCD147 CRC64;

Query Match 25.9%; Score 66.5; DB 11; Length 245;  
Best Local Similarity 33.3%; Pred. No. 0.6;  
Matches 18; Conservative 5; Mismatches 20; Indels 11; Gaps 2;

OY 1 TVSRVPTAMAFHGYRSTYOH-----DLRAYGFWRLL-----VRPVRHRRPVE 43  
 DB 180 TVSSGFWGVGYPGYRGYQYLLPEGDFRRHNMENWGAPOQMAVRRLRDRQWMOE 233

RESULT 2  
 ID Q9WVJ5 PRELIMINARY; PRT; 250 AA.  
 AC Q9WVJ5;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 GN BETA-BI CRYSTALLIN (3110006K12RIK PROTEIN).  
 DE CRYDBI OR 3110006K12RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96414434; PubMed=8817455;  
 RA Duncan M.K., Li X., Ogino H., Yasuda K., Piatigorsky J.;  
 RT "Developmental regulation of the chicken beta BI-crystallin promoter  
 in transgenic mice.";  
 RL Mech. Dev. 57:79-89(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Duncan M.K., Hejtmancik J.F., Piatigorsky J.;  
 RT "beta-BI crystallin: evidence for changes in gene regulation during  
 evolution.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi S., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staahl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarali R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamija M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombeerts P.,  
 RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 403:685-690(2001).  
 CC -1- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS OF  
 THE VERTEBRATE EYE LENS (BY SIMILARITY).  
 CC -1- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR VERY  
 CC SIMILAR GREEK KEY MOTIFS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BETA/GAMA-CRYSTALLIN FAMILY.  
 DR EMBL; AF106853; AAD2048.1; -;  
 DR HSSP; P02522; 2BB2.  
 DR MGD; MGI:104992; Ctrybhl.  
 DR InterPro; IPR001064; Crystallin.  
 DR Pfam; PF00030; crystall.2.  
 DR PRINTS; PR01367; BGCRCRYSTALLIN.  
 DR SMART; SM00247; XTALbg.2.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMA; 4.  
 KW Duplication: Eye lens protein.  
 SQ SEQUENCE 250 AA; 28002 MW; CCF26E7DD2F6B92EF CRC64;

Query Match 25.5%; Score 65.5; DB 11; Length 250;  
 Best Local Similarity 33.3%; Pred. No. 0.83;  
 Matches 18; Conservative 5; Mismatches 20; Indels 11; Gaps 2;

OY 1 TVSRVPTAMAFHGYRSTYOH-----DLRAYGFWRLL-----VRPVRHRRPVE 43  
 DB 185 TVSSGFWGVGYPGYRGYQYLLPEGDFRRHNMENWGAPOQMAVRRLRDRQWMOE 238

RESULT 3  
 ID Q9SD75 PRELIMINARY; PRT; 537 AA.  
 AC Q9SD75;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 GN HYPOTHETICAL 58.5 KDA PROTEIN.  
 DE F13112.30.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choisne N., Robert C., Brotlier P., Wincker P., Cattolico L.,  
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 DR EMBL; AL135292; CAB61944.1; -;  
 DR InterPro; IPR003662; sub-transporter.  
 DR Pfam; PF00083; sugar\_tr.1.  
 KW Hypothetical protein; Transmembrane.  
 SQ SEQUENCE 537 AA; 58503 MW; E4D85E54B9429A67 CRC64;

Query Match 23.9%; Score 61.5; DB 10; Length 537;  
 Best Local Similarity 45.5%; Pred. No. 6.3;  
 Matches 15; Conservative 3; Mismatches 6; Indels 9; Gaps 2;

OY 3 SRVPTAMAFHGYRSTYOHDLRAYGFWR-LLVRR 34  
 DB 363 SAVPWSMAFTGYIA-----GFWSDDLIR 387

RESULT 4  
 ID Q19842 PRELIMINARY; PRT; 724 AA.  
 AC Q19842;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 GN HYPOTHETICAL 79.8 KDA PROTEIN.  
 DE F27D9.5.  
 GN Caenorhabditis elegans.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Bentley D.;  
 RT "the sequence of *C. elegans* cosmid F27D9."  
 RL Submitted (Feb-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U49829; AAA93384.1; -.  
 DR HSSP; P24182; IBNC.  
 DR InterPro; IPR001249; AcCoA biotinCC.  
 DR InterPro; IPR001882; Biotin\_lipoyl.  
 DR InterPro; IPR000089; Biotin\_lipoyl.  
 DR InterPro; IPR000901; CPsase.  
 DR Pfam; PF02785; Biotin\_carb\_C; 1.  
 DR Pfam; PF00364; biotin\_lipoyl; 1.  
 DR Pfam; PF00289; CPsase\_L\_chain; 1.  
 DR Pfam; PF02786; CPsase\_L\_D2; 1.  
 DR PROSITE; PS00188; BIOTIN; UNKNOWN\_1.  
 DR PROSITE; PS00866; CPsase\_1; 1.  
 DR PROSITE; PS00867; CPsase\_2; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SO SEQUENCE 724 AA; 79762 MW; E09832FB65AABA45 CRC64;

Query Match 23.7%; Score 61; DB 5; Length 724;  
 Best Local Similarity 38.6%; Pred. No. 10;  
 Matches 17; Conservative 6; Mismatches 19; Indels 2; Gaps 2;

Oy 4 RVPWT-AMAFHGYR-STQHDLRAGFWRLVRRFVHRPHVESQ 45  
 Db 216 RVAMNDKQAREGYRLSKQEAASFQDDRMVLEKFDNPRHTEMQ 259

RESULT 5  
 ID 052214 PRELIMINARY; PRT; 288 AA.  
 AC 052214;  
 DT 01-NOV-1996 (TREMURel. 01, Created)  
 DT 01-NOV-1996 (TREMURel. 01, Last sequence update)  
 DE 01-JUN-2001 (TREMURel. 17, Last annotation update)  
 DE REPLICATION PROTEIN.  
 GN REPA.  
 OS *Yersinia enterocolitica*.  
 OG Plasmid pYv6227.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Yersinia*.  
 OX NCBI\_TaxID=630;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90299777; PubMed=1694522;  
 RA Vanocteghem J.C., Cornelis G.R.;  
 RT "Structural and functional similarities between the replication region  
 of the *Yersinia* virulence plasmid and the RepTIIA replicons."  
 RL J. Bacteriol. 172:3600-3608(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M22703;  
 RA Irlarte M., Lambermont I., Kerbouch C., Cornelis G.R.;  
 RT "Detailed genetic map of the pYv6227 plasmid of *Yersinia*  
*enterocolitica* serotype O:3."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; M55182; AAA98437.1; -.  
 DR EMBL; AF102990; AAD16866.1; -.  
 DR InterPro; IPR003446; IncFII\_repa.  
 DR Pfam; PF02387; IncFII\_repa; 1.  
 KW plasmid.  
 SO SEQUENCE 288 AA; 33548 MW; ABBFBA5FC74AAE45 CRC64;

Query Match 23.2%; Score 59.5; DB 2; Length 288;  
 Best Local Similarity 30.8%; Pred. No. 5.8;  
 Matches 20; Conservative 6; Mismatches 16; Indels 23; Gaps 3;

Oy 3 SRVFW-----TAMAF--HGYSYQHDLRAGFWRLVRRFVHR-R 39  
 Db 174 SRVEMNQOREKQRLPRLEMDLAKAMRFVRRFRSRYOTERKAHGKLRARARRDVRTR 233  
 Oy 40 PHVES 44  
 Db 234 RDIEA 238

RESULT 6  
 ID 09R3G9 PRELIMINARY; PRT; 288 AA.  
 AC 09R3G9;  
 DT 01-MAY-2000 (TREMURel. 13, Created)  
 DT 01-MAY-2000 (TREMURel. 13, Last sequence update)  
 DE 01-JUN-2001 (TREMURel. 17, Last annotation update)  
 DE PUTATIVE REPLICATION INITIATION PROTEIN.  
 GN YPCDL.77C OR REPA.  
 OS *Yersinia pestis*.  
 OG Plasmid pCD1.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Yersinia*.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 BIOVAR ORIENTALIS;  
 RA Baker S.G., Mungall K.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 BIOVAR ORIENTALIS;  
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 BIOVAR ORIENTALIS;  
 RA Karlyshev A.V., Wren B.W.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIM5;  
 RX MEDLINE=98427122; PubMed=9746557;  
 RA Perry R.D., Straley S.C., Featherston J.D., Rose D.J., Gregor J.,  
 RA Blattner F.R.;  
 RT "DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of  
*Yersinia pestis* KIM5."  
 RL Infect. Immun. 66:4611-4623(1998).  
 DR EMBL; AL117189; CAB54954.1; -.  
 DR EMBL; AF074612; AAC69762.1; -.  
 DR InterPro; IPR003446; IncFII\_repa.  
 DR Pfam; PF02387; IncFII\_repa; 1.  
 KW plasmid  
 SO SEQUENCE 288 AA; 33547 MW; 75D99C1DA548264D CRC64;

Query Match 23.2%; Score 59.5; DB 2; Length 288;  
 Best Local Similarity 30.8%; Pred. No. 5.8;  
 Matches 20; Conservative 6; Mismatches 16; Indels 23; Gaps 3;

Oy 3 SRVFW-----TAMAF--HGYSYQHDLRAGFWRLVRRFVHR-R 39  
 Db 174 SRVEMNQOREKQRLPRLEMDLAKAMRFVRRFRSRYOTERKAHGKLRARARRDVRTR 233  
 Oy 40 PHVES 44  
 Db 234 RDIEA 238

```

RESULT 7
068716 PRELIMINARY; PRT; 292 AA.
AC 068716;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
REPLICATION PROTEIN REPA HOMOLOG.
GN Yersinia pestis.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KIM.
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCready P., Skowronski E., Garnes J.,
RA Kobayashi A., Brubaker R.R., Garcia E.;
RT "Structural organization of virulence-associated plasmids of Yersinia
RT pestis.";
RT
RL J. Bacteriol. 180:5192-5202(1998).
DR EMBL: AF053946; AAC62601.1;
DR InterPro: IPR003446; Incefil_repa.
DR Pfam: PF02387; Incefil_repa; 1.
KW Plasmid.
SQ SEQUENCE 292 AA; 33989 MW; 1CE0B2E22C8EC074 CRC64;

Query Match
Best Local Similarity 23.2%; Score 59.5; DB 2; Length 292;
Matches 20; Conservative 6; Mismatches 16; Indels 23; Gaps 3;

QY 3 SHVPM-----TANAF--HGYSYQHDLRAYGFRLVRRFVHR-R 39
DB 178 SVENMNQOREKORLPRLMEDLIAMKRFVRRFRSYTERKAKHLKARARVDYDTR 237
QY 40 PVES 44
DB 238 RDIEA 242

RESULT 8
007420 PRELIMINARY; PRT; 322 AA.
AC 007420;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN HYPOTHETICAL 35.4 KDA PROTEIN.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Dayes R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL: 297050; CAB09743.1; -.
DR TubercuList; RV0176; -.

```

```

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 322 AA; 35437 MW; 1E32C3005970BED5 CRC64;

Query Match
Best Local Similarity 23.0%; Score 59; DB 16; Length 322;
Matches 15; Conservative 1; Mismatches 14; Indels 4; Gaps 1;

QY 8 TAW----AFHGYSYQHDLRAYGFRLVRRFVH 37
DB 83 TGMSLGRALTLGRVVRDSDSALGPMRLVRLDLAH 116

RESULT 9
09H495 PRELIMINARY; PRT; 336 AA.
AC 09H495;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ADIR2.
GN ADIR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Dron M., Merlet J.F., Tovey M.G.;
RT "Characterization of ADIR a novel Interferon responsive gene.";
RT Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ299441; CAC14461.1; -.
SQ SEQUENCE 336 AA; 38806 MW; 0053B78AF4FE967C CRC64;

Query Match
Best Local Similarity 22.8%; Score 58.5; DB 4; Length 336;
Matches 12; Conservative 5; Mismatches 8; Indels 5; Gaps 1;

QY 6 PWTANAFHGYSYQHDLRAYG-----FWRL 30
DB 36 PGSAMWPGFORLQEQLRAGALSKRYWL 65

RESULT 10
09H497 PRELIMINARY; PRT; 397 AA.
AC 09H497;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ADIR1.
GN ADIR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Dron M., Merlet J.F., Tovey M.G.;
RT "Characterization of ADIR1 a novel Interferon responsive gene.";
RT Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ299403; CAC13973.1; -.
DR InterPro: IPR001109; HuPF_Hypc.
SQ SEQUENCE 397 AA; 46139 MW; 459CD18EA22709FE CRC64;

Query Match
Best Local Similarity 22.8%; Score 58.5; DB 4; Length 397;
Matches 12; Conservative 5; Mismatches 8; Indels 5; Gaps 1;

QY 6 PWTANAFHGYSYQHDLRAYG-----FWRL 30

```

```

Db      36 PGSAMWPGFORLQEQRLAAGALSKRYWTL 65

RESULT 11
ID      09H6E7      PRELIMINARY;      PRT;      397 AA.
AC      09H6E7;
DT      01-MAR-2001 (TEMBLrel. 16, Created)
DT      01-MAR-2001 (TEMBLrel. 16, last sequence update)
DE      CDNA: FLJ22345 FIS, CLONE HRC06114 (ATP-DEPENDANT INTERFERON RESPONSE
        PROTEIN 1).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA      Suzuki Y., Ohtayashi M., Nishi T., Shibahara T., Tanaka T.,
RA      Nakamura Y., Isigaki T., Sugano S.;
RT      "NEO human cDNA sequencing project.";
RL      Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE-RHABDOMYOSARCOMA;
RA      Strausberg R.;
RN      Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE-EYE, RETINOBLASTOMA;
RA      Strausberg R.;
RN      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.
RC      TISSUE-MUSCLE, RHABDOMYOSARCOMA;
RA      Strausberg R.;
RL      Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AK025998; BAB15312.1; -.
DR      EMBL; BC001085; AAH01085.1; -.
DR      EMBL; BC011746; AAH11746.1; -.
DR      EMBL; BC007571; AAH07571.1; -.
DR      InterPro: IPR001109; HupE_HyPC.
SQ      SEQUENCE 397 AA; 46165 MW; 1E131DCCALAA6G31 CRC64;

Query Match      22.8%; Score 58.5; DB 4; Length 397;
Best Local Similarity 40.0%; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 8; Indels 5; Gaps 1;

QY      6 PMTAMAFHGRSYOHDLRAYG-----FWRL 30
Db      36 PGSAMWPGFORLQEQRLAAGALSKRYWTL 65

RESULT 12
ID      017732      PRELIMINARY;      PRT;      1175 AA.
AC      017732;
DT      01-JAN-1998 (TEMBLrel. 05, Created)
DT      01-JAN-1998 (TEMBLrel. 05, last sequence update)
DT      01-OCT-2001 (TEMBLrel. 18, last annotation update)
DE      PYRUVATE CARBOXYLASE.
GN      PYC-1 OR D2023.2.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC      Rhabditiidae; Pelodierinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Freedman J.H., Liao V.H.-C.;
RT      "Characterization of a cadmium-inducible form of pyruvate carboxylase
        from Caenorhabditis elegansl.";

```

```

RL      Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Kershaw J.;
RL      Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF237467; AAF60326.1; -.
DR      EMBL; Z81052; CAB02872.1; -.
DR      HSSP; P24182; 1BNC.
DR      InterPro: IPR000089; Biotin_lipoyl.
DR      InterPro: IPR000901; CPSase.
DR      InterPro: IPR000891; HMG_Like.
DR      InterPro: IPR003379; PYC_OADA.
DR      Pfam; PF02785; Biotin_carb_C; 1.
DR      Pfam; PF00364; Biotin_lipoyl; 1.
DR      Pfam; PF00289; CPSase_L_chain; 1.
DR      Pfam; PF02786; CPSase_L_D2; 1.
DR      Pfam; PF00682; HMG_Like; 1.
DR      Pfam; PF02436; PYC_OADA; 1.
DR      PROSITE; PS00867; CPSASE_2; UNKNOWN_1.
SQ      SEQUENCE 1175 AA; 129284 MW; DE3D21194C21E6AF CRC64;

Query Match      22.6%; Score 58; DB 5; Length 1175;
Best Local Similarity 43.3%; Pred. No. 42;
Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY      16 RSYOHDLRAYGFWRLRYRFVRRPHVESQ 45
Db      214 RSYSEAOAFGDCSLFYEKFEVERPRHIEVQ 243

RESULT 13
ID      052218      PRELIMINARY;      PRT;      290 AA.
AC      052218;
DT      01-NOV-1996 (TEMBLrel. 01, Created)
DT      01-NOV-1996 (TEMBLrel. 01, last sequence update)
DT      01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE      REPLICATION-ASSOCIATED PROTEIN.
GN      REPA.
OS      Klebsiella pneumoniae.
OG      Plasmid pIE545.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Klebsiella.
OX      NCBI_TaxID=573;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE-91177831; Pubmed=1706708;
RA      Praszler J., Wei T., Siemering K., Pittard J.;
RT      "Comparative analysis of the replication regions of IncB, IncK, and
        IncJ plasmids.";
RL      J. Bacteriol. 173:2393-2397(1991).
DR      EMBL; M93064; AAA98134.1; -.
DR      InterPro: IPR003446; IncFII_repa.
DR      Pfam; PF02387; IncFII_repa; 1.
KW      Plasmid.
SQ      SEQUENCE 290 AA; 33143 MW; 58C003F7E84A0571 CRC64;

Query Match      22.2%; Score 57; DB 2; Length 290;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 16; Conservative 5; Mismatches 11; Indels 22; Gaps 2;

QY      3 SRVPW-----TMAF--HGRSYOHDLRAYGFWRLLYR 34
Db      176 SRVEMENKLRKKGGLDGLMDELIRAMRFVRRFRSTQAEKSHGMKARARR 229

RESULT 14
ID      09WY09      PRELIMINARY;      PRT;      771 AA.
AC      09WY09;
DT      01-NOV-1999 (TEMBLrel. 12, Created)

```

DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE HYPOTHETICAL 87.5 KDA PROTEIN.  
 GN TM0162.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.L., Dodson R.J.,  
 Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RA "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima."  
 RL Nature 399:323-329(1999).  
 DR EMBL; AE001701; AAD35255.1; -.  
 DR TIGR; TM0162; -.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 771 AA; 87498 MW; EBB5D3781852A1D CRC64;

Query Match 22.2%; Score 57; DB 16; Length 771;  
 Best Local Similarity 36.7%; Pred. No. 36;  
 Matches 11; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

OY 15 YNSYQHDLRAYGFWRLLVRRFVRRPHEVS 44  
 DB 268 YSQYEPDILAHGTGLIVRLMRQHPHEVNN 297

RESULT 15  
 Q9DDT1 PRELIMINARY; PRT; 1180 AA.  
 AC Q9DDT1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE PYRUVATE CARBOXYLASE.  
 GN PC.  
 OS Brachydanio rerio (zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21100417; PubMed=11167010;  
 RA Yoder J.A., Litman G.W.;  
 RT "The zebrafish fchl, slc3a2, men1, pc, fgf3 and cycd1 genes define two  
 RT regions of conserved synteny between linkage group 7 and human  
 RT chromosome 11q13."  
 RL Gene 261:235-242(2000).  
 CC -1 - COFACTOR: BIOTIN (BY SIMILARITY).  
 DR EMBL; AF295372; AAG37836.1; -.  
 DR HSSP; P24182; 1BNC.  
 DR InterPro: IPR001882; Biotin.  
 DR InterPro: IPR000089; Biotin\_lipoyl.  
 DR InterPro: IPR000901; CPSase.  
 DR InterPro: IPR000891; HMGL-like.  
 DR InterPro: IPR003379; PYC\_OADA.  
 DR Pfam; PF02785; Biotin\_carb\_C; 1.  
 DR Pfam; PF00364; Biotin\_lipoyl; 1.  
 DR Pfam; PF00289; CPSase\_L\_chain; 1.  
 DR Pfam; PF02786; CPSase\_L\_D2; 1.  
 DR Pfam; PF00682; HMGL-like; 1.  
 DR Pfam; PF02436; PYC\_OADA; 1.  
 DR PROSITE; PS00188; BIOTIN; 1.

KW Biotin.  
 SQ SEQUENCE 1180 AA; 129884 MW; 26118F2691400E24 CRC64;

Query Match 22.2%; Score 57; DB 13; Length 1180;  
 Best Local Similarity 36.7%; Pred. No. 56;  
 Matches 11; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

OY 16 RSYQHDLRAYGFWRLLVRRFVRRPHEVSQ 45  
 DB 221 RAISEALAFNGALFVEKFEKPRHIEVQ 250

Search completed: August 15, 2002, 12:05:33  
 Job time: 373 sec





**THIS PAGE BLANK (USPTO)**

Fri Aug 16 10:36:46 2002

Seq. 105 (3-10) + 6 (3-10) + 7 (3-10)  
Residue residue residue

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:59:15 ; Search time 51.96 Seconds  
(Without alignments)  
51.304 Million cell updates/sec

Title: DEVI-613-COMBFrag  
Perfect score: 142  
Sequence: 1 SRPWTAWYQHDLRAYRRFVHRP 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A.Geneseq.032802.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	39.4	117	20	AAV45237 Mature KB5-C20 TCR
2	54	38.0	15	20	AAV30351 Epitope derived fr
3	53.5	37.7	96	22	AAU03032 Propionibacterium
4	51	35.9	512	22	ABB69433 Drosophila melanog
5	51	35.9	833	19	AAW42636 Protein sequence t
6	51	35.9	833	19	AAW42637 Protein sequence t
7	50	35.2	50	13	AAK1683 HSV-1 VP16 - prote
8	49	34.5	84	22	AAK77743 Human colon cancer
9	48.5	34.2	63	22	ABG22898 Novel human diagno
10	48.5	34.2	233	22	AAW39708 Human polypeptide
11	48.5	34.2	235	22	AAW41494 Human polypeptide

12	48	33.8	63	22	AAU49192 Propionibacterium
13	48	33.8	150	22	AAU42419 Propionibacterium
14	48	33.8	151	22	AAU41400 Propionibacterium
15	48	33.8	269	19	AAW70262 Amino acid sequenc
16	48	33.8	280	19	AAW86079 S. pneumoniae deri
17	48	33.8	280	19	AAW80697 S. pneumoniae deri
18	48	33.8	402	22	AAU45694 Propionibacterium
19	46.5	32.7	155	18	AAW10569 Mablin II prepro
20	46.5	32.7	155	18	AAW10569 Mablin II prepro
21	46	32.4	15	20	AAW30352 Epitope derived fr
22	46	32.4	43	20	AAW25799 Human secreted pro
23	46	32.4	62	22	ABG16003 Novel human diagno
24	46	32.4	247	22	AAU50023 Propionibacterium
25	46	32.4	396	20	AAV25793 Human secreted pro
26	46	32.4	460	22	ABG06687 Novel human diagno
27	46	32.4	781	22	ABG14190 Novel human diagno
28	45.5	32.0	282	22	ABB66530 Drosophila melanog
29	45.5	32.0	378	22	ABB66530 Drosophila melanog
30	45	31.7	15	20	AAV30353 Epitope derived fr
31	45	31.7	50	22	AAU42356 Propionibacterium
32	45	31.7	182	21	AAK33129 Zea mays protein f
33	45	31.7	451	16	AAW79909 N. crassa mtr gene
34	45	31.7	470	15	AAW47118 Mtr protein of Neu
35	45	31.7	488	20	AAV05830 Human altered sph
36	45	31.7	568	20	AAV15211 sphingosine-1-phos
37	45	31.7	568	20	AAV05827 Human sphingosine-
38	45	31.7	568	22	AAW78463 Human protein SEQ
39	45	31.7	580	22	AAW79445 Human protein SEQ
40	45	31.7	754	21	AAW11529 SEN virus protein
41	44	31.0	62	22	AAW80330 Human haematologic
42	44	31.0	124	22	ABW11127 Human gal-1-P urid
43	44	31.0	158	22	AAU48192 Propionibacterium
44	44	31.0	431	22	ABW1406 Human N-copine hom
45	44	31.0	461	22	AAU35244 Enterococcus faeca

ALIGNMENTS

RESULT 1  
AAV45237 standard; peptide: 117 AA.  
XX  
XX AAV45237:  
XX  
XX 06-JAN-2000 (first entry)  
XX  
XX  
DE Mature KB5-C20 TCR scfv:Valpha-linker-Vbeta Vbeta peptide.  
XX  
XX scfv: TCR: V domain: T cell receptor; Immune system; Infection;  
KW tumour; Immune response; linker.  
XX  
XX OS Synthetic.  
XX  
XX CA2205881-A.  
XX  
XX  
XX 09-JAN-1999.  
XX  
XX  
XX 09-JUL-1997; 97CA-2205881.  
XX  
XX 09-JUL-1997; 97CA-2205881.  
XX  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX Mazza G, Malissen B, Gregoire C;  
XX  
XX WPI: 1999-580982/50.  
XX  
XX New T cell receptor peptides, useful for treating and/or preventing  
XX myelomas and bacterial, viral and fungal infections -  
XX  
XX Example 1; Fig 1B; 84pp; English.  
XX

Query Match	39.4%	Score 56;	DB 20;	Length 117;
Best Local Similarity	57.1%;	Pred No. 1,2;		
Matches	8;	Conservative	3;	Mismatches 3; Indels 0; Gaps 0.
OY	1	SRPWTANYQHDR	14	
		1: 1: 1: 1: 1: 1:		
db	28	sqypmmswygqdlq	41	

RESULT 2  
 ID AAY30351 standard; Peptide: 15 AA.  
 XX  
 AC AAY30351;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 DE Epitope derived from pneumococcal surface adhesion A protein.  
 XX  
 KW Pneumococcal surface adhesion A protein; Psaa: monoclonal antibody;  
 KW vaccine; Streptococcus pneumoniae infection.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN MO9945121-A1.  
 XX  
 PD 10-SEP-1999.  
 XX  
 PE 26-FEB-1999; 99WO-US04326.  
 XX  
 PR 02-MAR-1998; 98US-0076565.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Adès EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;  
 PI Zeller JL;  
 XX  
 XR WPI: 1999-540849/45.  
 XX  
 PT New peptides corresponding to Streptococcus pneumoniae Psaa, used  
 PT for treating or preventing Streptococcus pneumoniae infection in a  
 PT subject  
 XX  
 PS Claim 6; Page 43; 58pp; English.  
 XX  
 CC AAY30351-54 represent immunogenic peptides which are derived from  
 CC a pneumococcal surface adhesion A protein (psaa). The specification  
 CC describes monoclonal antibodies which bind epitopes of the psaa protein

CC (e.g. present sequence). The peptides can be used in vaccines to prevent  
CC Streptococcus pneumoniae infections. The antibodies of the invention  
CC can also be used to detect *S. pneumoniae* in a sample or individual.  
XX  
SQ Sequence 15 AA;

Query Match	38.0%;	Score 54;	DB 20;	Length 15;
Best Local Similarity	81.8%;	Pred. No. 0.26;		
Matches	9;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0;

```
QY      1 SRVPWTAWYQH 11
          |||||
Db      3 srvpwtawafh 13
```

RESULT	3
AAU40302	
ID	AAU40302 standard; Protein; 96 AA.
XX	
AC	AAU40302;
XX	
DT	13-FEB-2002 (first entry)

KM SAPHO syndrome; synovitis; acne; pustulosis; hyperkeratosis; osteomyelitis  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ERISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuropsectant.

**OS** **Propionibacterium acnes.**

PN W0200181581-A2.

PD 01-NOV-2001

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatla A;

XX

DR N-PSDB; AAS59511.

PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT treating acne vulgaris -

PS Example 1; SEQ ID NO 1497; 1069pp; English

CC Sequences MANU319105-MAN68017 represent propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAMPO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 96 AA;

Query Match 37.7%; Score 53.5; DB 22; Length 96;  
 Best Local Similarity 46.2%; Pred. No. 2.2;

Matches 12; Conservative 1; Mismatches 10; Indels 3; Gaps 1;

OY 2 RVPTAWYOHDLRAYRRFV---HRRP 24  
 | : | | | | | | | | | |  
 Db 44 rlprrflwphrdirdiramvaahrtr 69

RESULT 4

ABB69433  
 ID ABB69433 standard; Protein; 512 AA.

XX ABB69433:

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 35091.

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.

XX Drosophila melanogaster.

XX MO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PDB; ABL13536.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

XX Disclosure: SEQ ID NO 35091; 21pp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB161840-AB16175) and the encoded proteins  
 CC (AB161737-AB161702).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 512 AA;

Query Match 35.9%; Score 51; DB 22; Length 512;  
 Best Local Similarity 50.0%; Pred. No. 29;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 7 AMYOHDLRAYRRFVHRRP 24  
 : | | | | | : | |  
 Db 340 syfhdlnrlyrtlsqr 357

RESULT 5

AAW42636  
 ID AAW42636 standard; Protein; 833 AA.

XX AAW42636:

XX 22-OCT-1998 (first entry)

DE Protein sequence that is specific for *Neisseria meningitidis*.

KW *N. gonorrhoeae*; *N. lactamica*; chromosome 22491; region 1; region 2;  
 KM region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
 KM meningitis.

XX *Neisseria meningitidis*.

XX WO9802547-A2.

XX 22-JAN-1998.

XX 11-JUL-1997; 97WO-FR01295.

XX 12-JUL-1996; 96FR-0008768.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX (SMIK ) SMITHKLINE BEECHAM.

XX Nassif X, Tinsley C, Achtman M, Werker P, Ruelle J;

XX Vinals C;

XX WPI; 1998-110594/10.

XX N-PDB; AAW03553.

XX Claim 8; Pages 91-95; 150pp; French.

XX AAW42633-37 and AAW42639-41 are encoded by a DNA sequence found in region  
 CC 2 of *Neisseria meningitidis*. The specification describes DNA sequences  
 CC that are found in *N. meningitidis*, but not in *N. gonorrhoeae* or  
 CC *N. lactamica*, except for the genes involved in biosynthesis of the  
 CC capsule polysaccharide, frpA or C, opc, porA, rotamase, sequence IC1106,  
 CC IGA protease, pillin, pilC, proteins which bind transferrin and opacity  
 CC proteins. The DNA sequences are found on chromosome 22491, mainly (or  
 CC within 20 kb) between tufa and pilT (region 1), pilQ and lambda-740  
 CC (region 2) or argF and opaB (region 3). The DNA sequences are  
 CC responsible for the differences in pathogenicity between *N. meningitidis*  
 CC and *N. gonorrhoeae*, specifically they include the genes that allow  
 CC *N. meningitidis* to cross the blood-brain barrier. DNA sequences common  
 CC to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N. lactamica*, are  
 CC responsible for colonisation and penetration of the mucosa. The DNA  
 CC sequences can be used to produce probes and primers, and antibodies  
 CC produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.

XX Sequence 833 AA;

Query Match 35.9%; Score 51; DB 19; Length 833;  
 Best Local Similarity 33.3%; Pred. No. 48;  
 Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

OY 2 RVPTAWYOHDLRAYRRFVHR 22



KW colorectal carcinoma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200122920-A2.  
 XX  
 PD 05-APR-2001.  
 XX  
 PF 28-SEP-2000; 2000MO-US26524.  
 XX  
 PR 29-SEP-1999; 99US-0157137.  
 PR 03-NOV-1999; 99US-0163280.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Barash SC, Birse CE, Rosen CA.  
 XX  
 DR WPI; 2001-235357/24.  
 DR N-PSDB; AAH37150.  
 XX  
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 XX  
 XX Claim 11: Page 9762-9763; 9803pp; English.  
 XX  
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patient's own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated Ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAB77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 CC  
 XX  
 SO Sequence 84 AA;  
 XX  
 Query Match 34.5%; Score 49; DB 22; Length 84;  
 Best Local Similarity 63.6%; Pred. No. 8.4;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 OY 4 PWTAWYOHDLR 14  
 | | | | | | : : |  
 Db 63 pwtawxxhnr 73  
 XX  
 RESULT 9  
 ABG22898  
 ID ABG22898 standard; Protein; 63 AA.  
 XX  
 AC ABG22898;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #22889.  
 XX  
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.

XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001MO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS87085.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20: SEQ ID NO 53257; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX  
 SO Sequence 63 AA;  
 XX  
 Query Match 34.2%; Score 48.5; DB 22; Length 63;  
 Best Local Similarity 45.8%; Pred. No. 7.4;  
 Matches 11; Conservative 1; Mismatches 9; Indels 3; Gaps 1;  
 OY 1 SRVPWTAWYOHDLRAVRRFVHRRP 24  
 | | | | | | | | | | | |  
 Db 25 srpwwksvtrpdrgrar---trp 45  
 XX  
 RESULT 10  
 AAM39708  
 ID AAM39708 standard; Protein; 233 AA.  
 XX  
 AC AAM39708;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 2853.  
 XX  
 XX Human: neotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX

```

OS Homo sapiens.
XX
XX WO200153312-A1.
XX
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB: AA158864.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2853; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AA158642-AA162213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 233 AA:
SQ

```

```

Query Match 34.2%; Score 48.5; DB 22; Length 233;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 9; Conservative 1; Mismatches 7; Indels 1; Gaps 1;
OY 4 PWTAWYQHDLRAYRRFVH 21
   || || || || ||
Db 120 pWLVWYSR-YRAWREGVH 136

```

```

RESULT 11
AA161494
ID AA161494 standard; Protein: 235 AA.
XX
XX AA161494;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 6425.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW

```

```

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB: AA160650.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6425; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AA158642-AA162213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 235 AA:
SQ

```

```

Query Match 34.2%; Score 48.5; DB 22; Length 235;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 9; Conservative 1; Mismatches 7; Indels 1; Gaps 1;
OY 4 PWTAWYQHDLRAYRRFVH 21
   || || || || ||
Db 122 pWLVWYSR-YRAWREGVH 138

```

```

RESULT 12
AA161912
ID AA161912 standard; Protein: 63 AA.
XX
XX AA161912;
XX
XX 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #10088.
XX
XX

```



```

XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN W0200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001MO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59545.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 10387; 10699p; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 63 AA;

```

Query Match 33.8%; Score 48; DB 22; Length 63;

Best Local Similarity 50.0%; Pred. No. 8.7; Mismatches 8; Indels 0; Gaps 0;

```

Qy 2 RVPWTAWYQHDLRAYRRFVH 21
   1 1 1 1 1 1 1 1 1 1 1
Db 24 rvpqtahwchrrrcrrffh 43

```

# RESULT 13

AAU42419 standard; Protein: 150 AA.

AAU42419;

27-FEB-2002 (first entry)

DT XX

```

DE Propionibacterium acnes immunogenic protein #3315.
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN W0200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001MO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59517.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 3614; 10699p; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 150 AA;

```

Query Match 33.8%; Score 48; DB 22; Length 150;

Best Local Similarity 50.0%; Pred. No. 22; Mismatches 10; Indels 0; Gaps 0;

```

Qy 3 VPMWTAWYQHDLRAYRRFVHRP 24
   1 1 1 1 1 1 1 1 1 1 1
Db 30 vpggstaahdirlyflgdvthp 51

```

# RESULT 14

AAU41400 standard; Protein: 151 AA.

AAU41400;

13-FEB-2002 (first entry)

DT XX

```
XX DE Propionibacterium acnes immunogenic protein #2296.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopahlic; neuroprotectant.
XX OS
XX PN Propionibacterium acnes.
XX PD WO200181581-A2.
XX PD 01-NOV-2001.
XX PE 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonmeuve J, Zhang Y, Jen S, Carter D;
XX DR WPI: 2001-616774/71.
XX DR N-PSDB; AAS59515.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX PS
XX PS Example 1; SEQ ID NO 2595; 1069pp; English.
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX SQ Sequence 151 AA:

Query Match 33.8%; Score 48; DB 22; Length 151;
Best Local Similarity 37.0%; Pred. No. 22;
Matches 10; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

QY 2 RVPWTAWYQ---HDLRAYRFVHRRP 24
   | | | | | | | | | | | | |
Db 59 rpfqgwhrvhvgprdrprstrhrhrcp 85

RESULT 15
AAW70262
ID AAW70262 standard; Protein: 269 AA.
XX AC AAW70262;
XX
```

```
DT 20-NOV-1998 (first entry)
XX DE Amino acid sequence of Streptococcus pneumoniae l1cd2 protein.
XX KW l1cd2 gene; antagonist; inhibition; antibody; T cell; immune response;
XX KW vaccination; otitis media; conjunctivitis; pneumonia; bacteraemia;
XX KW pleural empyaema; meningitis; sinusitis; endocarditis.
XX OS Streptococcus pneumoniae.
XX PN EP861889-A1.
XX PD 02-SEP-1998.
XX PD 27-FEB-1998; 98EP-0301446.
XX PF 28-FEB-1997; 97US-0039225.
XX PR (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA Lonetto MA;
XX PI WPI: 1998-449106/39.
XX DR N-PSDB; AAV33008.
XX PT New DNA encoding Streptococcus pneumoniae l1cd2 polypeptides - used
XX PT to treat, prevent, diagnose and vaccinate against e.g. otitis media,
XX PT conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis,
XX PT pleural empyaema and endocarditis
XX PS
XX PS Claim 1; Page 5; 29pp; English.
XX CC This is the amino acid sequence of the Streptococcus pneumoniae l1cd2
XX CC protein, a choline kinase used in the method of an individual in need of
XX CC l1cd2 polypeptide is used for the treatment of an individual in need of
XX CC l1cd2 protein and the antagonist is used for the treatment of an
XX CC individual with the need to inhibit l1cd2 polypeptide. Inducing an
XX CC immunological response in a mammal comprises inoculating the mammal with
XX CC the l1cd2 polypeptide to produce antibody and/or T cell immune responses.
XX CC l1cd2 is used to prevent, treat, diagnose and vaccinate against e.g.
XX CC otitis media, conjunctivitis, pneumonia, bacteraemia, pleural empyaema,
XX CC meningitis, sinusitis and endocarditis.
XX SQ
XX SQ Sequence 269 AA:

Query Match 33.8%; Score 48; DB 19; Length 269;
Best Local Similarity 54.5%; Pred. No. 40;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 WYQHDRLAYRR 18
   | | | | | | |
Db 258 wyshtskayrk 268
```

Search completed: August 15, 2002, 11:59:16  
Job time: 531 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 11:59:42 ; Search time 20.64 Seconds  
(Without alignments)  
28.402 Million cell updates/sec

Title: DEVI-613-COMBFRAG  
Perfect score: 142  
Sequence: 1 SRVPTAWYQHDLRAYRFRVHRRP 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46.5	32.7	155	2	US-08-670-186-2
2	45	31.7	451	1	US-08-191-337-3
3	45	31.7	568	4	US-09-238-373-2
4	44	31.0	669	2	US-07-861-800-2
5	44	31.0	2465	2	US-08-596-291-3
6	44	31.0	2465	3	US-09-100-804-3
7	44	31.0	2466	3	US-09-080-855-12
8	44	31.0	2466	5	PCT-US94-09943-2
9	44	31.0	2485	4	US-09-290-640-46
10	43	30.3	310	2	US-08-640-765A-1
11	43	30.3	310	4	US-09-073-613-1
12	43	30.3	1257	4	US-08-340-428B-49
13	43	30.3	1312	4	US-08-989-299-8
14	42	29.6	17	1	US-08-465-91A-387
15	42	29.6	17	1	US-08-464-538B-384
16	42	29.6	17	2	US-08-463-076E-378
17	42	29.6	770	1	US-08-369-796-12
18	42	29.6	770	1	US-08-416-581B-5
19	42	29.6	770	1	US-08-416-581B-5
20	42	29.6	770	1	US-08-416-581B-5
21	42	29.6	770	2	US-08-852-091-12
22	42	29.6	770	2	US-08-820-754-12
23	42	29.6	770	3	US-08-956-652-12
24	42	29.6	770	3	US-08-956-869-12
25	42	29.6	770	3	US-09-012-710-8
26	42	29.6	770	3	US-08-948-547-12
27	42	29.6	770	4	US-09-087-465-6

28	42	29.6	770	4	US-09-364-970-3	Sequence 3, Appl1
29	42	29.6	770	4	US-09-364-970-5	Sequence 5, Appl1
30	42	29.6	770	4	US-09-556-273-8	Sequence 8, Appl1
31	42	29.6	770	4	US-08-956-653A-12	Sequence 12, Appl1
32	42	29.6	770	5	PCT-US95-17025-12	Sequence 12, Appl1
33	42	29.6	771	1	US-08-276-099A-14	Sequence 14, Appl1
34	42	29.6	771	1	US-08-781-890-14	Sequence 14, Appl1
35	42	29.6	1310	4	US-08-989-299-10	Sequence 10, Appl1
36	41	28.9	67	2	US-08-248-839C-100	Sequence 100, App
37	41	28.9	781	1	US-08-280-690-2	Sequence 2, Appl1
38	41	28.9	2409	6	5180808-2	Sequence 2, Appl1
39	40.5	28.5	471	4	US-09-160-494-6	Sequence 6, Appl1
40	40	28.2	148	2	US-09-193-877-1	Sequence 1, Appl1
41	40	28.2	260	2	US-09-193-877-3	Sequence 3, Appl1
42	40	28.2	294	1	US-08-325-562-2	Sequence 2, Appl1
43	40	28.2	294	1	US-08-437-795-2	Sequence 2, Appl1
44	40	28.2	318	4	US-09-199-637A-23	Sequence 23, Appl1
45	40	28.2	471	4	US-09-071-709-4	Sequence 4, Appl1

## ALIGNMENTS

RESULT 1  
US-08-670-186-2  
; Sequence 2, Application US/08670186  
; Patent No. 5859343  
GENERAL INFORMATION:  
; APPLICANT: SUN, SAMUEL S. M.  
; APPLICANT: XIONG, LIMIN  
; APPLICANT: HU, ZHONG  
; APPLICANT: CHEN, HANG  
; TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/670,186  
; FILING DATE: 21-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MORASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 23461-20007.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 822-0168  
; TELEX: 90-4030 MRSNROERSMH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-670-186-2

Query Match 32.7%; Score 46.5; DB 2; Length 155;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
OY 10 QHD-LRAYRFRVHRR 23  
|| ||| :||:||||

Db 46 OHORLACORFTHRR 60

RESULT 2

US-08-191-337-3

; Sequence 3, Application US/08191337  
; Patent No. 5643745

; GENERAL INFORMATION:

; APPLICANT: STUART, W. DORSEY

; TITLE OF INVENTION: HETEROLOGOUS DIMERIC PROTEINS PRODUCED

; TITLE OF INVENTION: IN HETEROKARYONS

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 2000 Pennsylvania Ave. N.W.

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20006-1812

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/191,337

; FILING DATE: 03-FEB-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: MORASHIGE, KATE H.

; REGISTRATION NUMBER: 29,959

; REFERENCE/DOCKET NUMBER: 3918-0002.00

; TELEPHONE: (202) 887-1500

; TELEFAX: (202) 887-0763

; TELEX: 90-4030

; INFORMATION FOR SEQ. ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 451 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-191-337-3

Query Match

Best Local Similarity 31.7%; Score 45; DB 1; Length 451;

Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 VPTWTMYOHDLRAYRRFV 20

Db 227 VPMSCWPKEDLSLAEGFI 244

RESULT 3

US-09-238-373-2

; Sequence 2, Application US/09238373A

; Patent No. 6187562

; GENERAL INFORMATION:

; APPLICANT: DUCKWORTH, DAVID MALCOLM

; APPLICANT: GODDEN, ROBERT JAMES

; APPLICANT: TESTA, TANITA TAMSON

; TITLE OF INVENTION: NOVEL COMPOUNDS

; FILE REFERENCE: GP-30034

; CURRENT APPLICATION NUMBER: US/09/238,373A

; CURRENT FILING DATE: 1999-01-27

; EARLIER FILING DATE: 1998-11-03 9824026.0

; EARLIER APPLICATION NUMBER: EP 98300625.5

; EARLIER FILING DATE: 1998-01-29

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 2

; LENGTH: 568

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-238-373-2

Query Match

Best Local Similarity 31.7%; Score 45; DB 4; Length 568;

Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 RVPWTMYOHDLRAYRRFVR 22

Db 258 RVPWTMYOHDLRAYRRFVR 278

RESULT 4

US-07-861-800-2

; Sequence 2, Application US/07861800

; Patent No. 5925515

; GENERAL INFORMATION:

; APPLICANT: VAN HARTINGSVELDT, WILLEM

; APPLICANT: VAN GORCOM, ROBERT

; APPLICANT: GOUKA, R. J.

; APPLICANT: BOVENBERG, ROELOF A. L.

; TITLE OF INVENTION: Transformant Selection Marker System

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/861,800

; FILING DATE: 19920611

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: MORASHIGE, KATE H.

; REGISTRATION NUMBER: 29,959

; REFERENCE/DOCKET NUMBER: 44615-20018.20

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; TELEX: 706141

; INFORMATION FOR SEQ. ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 669 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-861-800-2

Query Match

Best Local Similarity 31.0%; Score 44; DB 2; Length 669;

Matches 7; Conservative 6; Mismatches 7; Indels 4; Gaps 1;

QY 1 SRVPWTA---WYOHDLRAYRRFV 20

Db 243 AEVPTWAGRDIMWHHEVEKYPNVL 266

RESULT 5

US-08-596-291-3

; Sequence 3, Application US/08596291

; Patent No. 5821075

; GENERAL INFORMATION:

; APPLICANT: GONZALEZ, LEONEL JORGE

; APPLICANT: SARAS, JAN

```

1  APPLICANT:  CLASSON-WELSH, LENA
2  APPLICANT:  HELDIN, CARL-HENRIK
3  TITLE OF INVENTION:  PRIMARY STRUCTURE AND FUNCTIONAL
4  TITLE OF INVENTION:  EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
5  TITLE OF INVENTION:  TYROSINE PHOSPHATASES
6  NUMBER OF SEQUENCES:  4
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE:  WOLF, GREENFIELD & SACKS, P.C.
9  STREET:  600 ATLANTIC AVENUE
10 CITY:  BOSTON
11 STATE:  MASSACHUSETTS
12 COUNTRY:  USA
13 ZIP:  02210
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE:  Floppy disk
17
18 MEDIUM TYPE:  IBM PC compatible
19 OPERATING SYSTEM:  PC-DOS/MS-DOS
20 SOFTWARE:  Patent In Release #1.0, Version #1.25
21
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER:  US/08/596,291
24 FILING DATE:  09-AUG-1996
25 CLASSIFICATION:  435
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER:  US 08/115,573
28 FILING DATE:  01-SEP-1993
29 ATTORNEY/AGENT INFORMATION:
30 NAME:  GATES, EDWARD R.
31 REGISTRATION NUMBER:  31,616
32 REFERENCE/DOCKET NUMBER:  LC0461/7000
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE:  617/720-3500
35 TELEFAX:  617/720-2441
36
37 TELEX:  92-1742 EZEKTEL
38 INFORMATION FOR SEQ ID NO:  3:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH:  2465 amino acids
41 TYPE:  amino acid
42 TOPOLOGY:  linear
43 MOLECULE TYPE:  protein
44
45 US-08-596-291-3
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
93
```

```

COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: L0461/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2465 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-100-804-3

Query Match          31.0%; Score 44; DB 3; Length 2465;
Best Local Similarity 31.2%; Pred. No. 5.6e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

OY 1 SRVPTANYQHD-----LRAVREYVR 22
      | : ||| || : | : ||
Db 2349 SHLNETAMPDHDPSOPDDLTFISYMRHHR 2380

RESULT 7
US-09-080-855-12
; Sequence 12, Application US/09080855A
; Patent No. 6083721
; GENERAL INFORMATION:
; APPLICANT: Saras, Jan
; APPLICANT: Franz, Petra
; APPLICANT: Aspenstrm, Pontus
; APPLICANT: Hellman, Ulf
; APPLICANT: Gomez, Leonel Jorge
; APPLICANT: Heidln, Carl-Henrik
; TITLE OF INVENTION: PARC: A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PP1L
; FILE REFERENCE: L0461/7030
; CURRENT APPLICATION NUMBER: US/09/080,855A
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: 08/805,583
; EARLIER FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-080-855-12

Query Match          31.0%; Score 44; DB 3; Length 2466;
Best Local Similarity 31.2%; Pred. No. 5.6e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

OY 1 SRVPTANYQHD-----LRAVREYVR 22
      | : ||| || : | : ||

```

```

Db      2350  SHLNFTAMPDHPDPSQPDLLFTFSYMRHHR 2381

RESULT      8
PCT-US94-09943-2
; Sequence 2, Application PC/TUS9409943
; GENERAL INFORMATION:
;   APPLICANT:
;   APPLICANT:
;   APPLICANT:
;   APPLICANT:
;   APPLICANT:
;   TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
;   TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   STREET: 600 ATLANTIC AVENUE
;   CITY: BOSTON
;   STATE: MASSACHUSETTS
;   COUNTRY: USA
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US94/09943
;   FILING DATE: 01-SEP-1994
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/115,573
;   FILING DATE: 01-SEP-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: TWOMEY, MICHAEL J.
;   REGISTRATION NUMBER: P-38,349
;   REFERENCE/DOCKET NUMBER: 10461/7000WO
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/720-3500
;   TELEFAX: 617/720-2441
;   TELEX: 92-1742 EZEKIEL
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2466 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   PCT-US94-09943-2

Query Match      31.0%; Score 44; DB 5; Length 2466;
Best Local Similarity 31.2%; Pred. No. 5.6e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 10; Gaps 1

QY      1  SRYPWTAWYQHD-----LRAVRRFVHR 22
      I : : : : I : : : : I : : : :
Db      2350  SHLNFTAMPDHPDPSQPDLLFTFSYMRHHR 2381

RESULT      9
US-09-290-640-46
; Sequence 46, Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
;   APPLICANT: Dean, Nicholas M.
;   APPLICANT: Marcusson, Eric G.
;   TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
;   FILE REFERENCE: ISPh-Q051
;   CURRENT APPLICATION NUMBER: US/09/290,640
;   CURRENT FILING DATE: 1999-04-12
;   NUMBER OF SEQ ID NOS: 85
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 46

```

```

: LENGTH: 2485
: TYPE: PROT
: ORGANISM: Homo sapiens
US-09-290-640-46

Query Match
Best Local Similarity 31.0%; Score 44; DB 4; Length 2485;
Matches 10; Conservative 4; Mismatches 8; Indels 10; Gaps 1

QY 1 SRVFWTAWYQHD-----LRAYRRFVR 22
      | : || | | : | | | | : | | |
Db 2369 SHLNFTAMPDHDTPSQPDDLTFISYMRHHR 2400

RESULT 10
US-08-640-765A-1
: Sequence 1, Application US/08640765A
: Patent No. 5821103
: GENERAL INFORMATION:
: APPLICANT: Tanuma, Sei-ichi
: TITLE OF INVENTION: NOVEL DEOXYRIBONUCLEASE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Leydig, Voit & Mayer
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/640,765A
FILING DATE: 06-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6239518
FILING DATE: 06-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Green, Robert F.
REGISTRATION NUMBER: 27555
REFERENCE/DOCKET NUMBER: 73362
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-640-765A-1

Query Match
Best Local Similarity 30.3%; Score 43; DB 2; Length 310;
Matches 10; Conservative 3; Mismatches 11; Indels 4; Gaps 1

QY 1 SRVFWTAWYQHDLRAYRRFV---HRRP 24
      | | | : | | | : | | | | |
Db 136 SREPFVWFQAPFTAAKDFVIVPLHTTP 163

RESULT 11
US-09-073-613-1
: Sequence 1, Application US/09073613
: Patent No. 6143875

```



```

: GENERAL INFORMATION:
: APPLICANT: Tanuma, Sei-ichi
: TITLE OF INVENTION: ANTIBODY TO NOVEL DEOXYRIBONUCLEASE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Leydig, Voigt & Mayer
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/073,613
: FILING DATE: 06-MAY-1998
: CLASSIFICATION:
: APPLICATION NUMBER: 08/640,765
: FILING DATE:
: APPLICATION NUMBER: JP 6239518
: FILING DATE: 06-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: LARCHER, Carol
: REGISTRATION NUMBER: 35243
: REFERENCE/DOCKET NUMBER: 85519
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 616-5600
: TELEFAX: (312) 616-5700
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 310 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-073-613-1

Query Match      30.3%; Score 43; DB 4; Length 310;
Best Local Similarity 35.7%; Pred. No. 98;
Matches 10; Conservative 3; Mismatches 11; Indels 4; Gaps 1;

QY      1 SRVPTAWIOHDLRAYRREV----HRRP 24
DB      136 SRPFVWFQAPFTAKDFIVPLHTTP 163

RESULT 12
US-08-340-428B-49
: Sequence 49, Application US/08340428B
: Patent No. 5648465
: GENERAL INFORMATION:
: APPLICANT: MARGOLIS, Richard U.
: APPLICANT: RACH, Uwe
: APPLICANT: MARGOLIS, Renee K.
: TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
: TITLE OF INVENTION: CLEUOCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
: NUMBER OF SEQUENCES: 49
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Browdy and Neimark
: STREET: 419 Seventh Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/340,428B
: FILING DATE: 14 NO. 5648465ember 1994
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/922,911
: FILING DATE: 03 August 1992
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Browdy, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: Margolis-1A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 49:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1257 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-340-428B-49

Query Match      30.3%; Score 43; DB 1; Length 1257;
Best Local Similarity 43.8%; Pred. No. 4e+02;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY      8 WYOHDLRAYRFFVHRR 23
DB      1033 WHKFGCHCYRFFAHR 1048

RESULT 13
US-08-989-299-8
: Sequence 8, Application US/08989299
: Patent No. 6194556
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan L.
: APPLICANT: Robinson, Keith E.
: TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
: TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: FOLEY, HOAG & ELIOT LLP
: STREET: One Post Office Square
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109-2170
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/989,299
: FILING DATE: 11-DEC-1997
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Arnold E., Beth
: REGISTRATION NUMBER: 35,430
: REFERENCE/DOCKET NUMBER: MIA-025.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-832-1000
: TELEFAX: 617-832-7000
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1312 amino acids
: TYPE: amino acid
: STRANDEDNESS:

```

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-989-299-8

Query Match 30.3%; Score 43; DB 4; Length 1312;  
Best Local Similarity 28.2%; Pred. No. 4.1e+02;  
Matches 11; Conservative 4; Mismatches 4; Indels 20; Gaps 2;

OY 5 TWAMYQH-----DLRAY-RPFVHRR 23  
DB 232 WRSWSPSPSESLBHYHQLEPLVNLHAYVRRALHRR 270

RESULT 14  
US-08-465-391A-387  
Sequence 387, Application US/08465391A  
Patent No. 5786331  
GENERAL INFORMATION:  
APPLICANT: Barrett, Ronald W.  
APPLICANT: Yanofsky, Stephen D.  
APPLICANT: Baldwin, David  
APPLICANT: Jacobs, Jeff W.  
APPLICANT: Boy, Philippe R.  
APPLICANT: Leahy, Ellen M.  
APPLICANT: Pottorf, Richard S.  
TITLE OF INVENTION: Peptides and Compounds That Bind to the  
NUMBER OF SEQUENCES: 405  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465.391A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/373.474  
FILING DATE: 01-FEB-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/190.788  
FILING DATE: 02-FEB-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5786331viel, Vern  
REGISTRATION NUMBER: 32,483  
REFERENCE/DOCKET NUMBER: 16528A-001840/1019.2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 387:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-465-391A-387

Query Match 29.6%; Score 42; DB 1; Length 17;  
Best Local Similarity 63.6%; Pred. No. 7.5;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 6 TAWYQHDRLAY 16  
DB 7 TAWYDHFLLSY 17

RESULT 15  
US-08-464-538B-384  
Sequence 384, Application US/08464538B  
Patent No. 5861476  
GENERAL INFORMATION:  
APPLICANT: Barrett, Ronald W.  
APPLICANT: Yanofsky, Stephen D.  
APPLICANT: Baldwin, David  
APPLICANT: Jacobs, Jeff W.  
APPLICANT: Boy, Philippe R.  
APPLICANT: Leahy, Ellen M.  
APPLICANT: Pottorf, Richard S.  
TITLE OF INVENTION: Peptides and Compounds That Bind to the  
NUMBER OF SEQUENCES: 402  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464.538B  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/373.474  
FILING DATE: 01-FEB-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/190.788  
FILING DATE: 02-FEB-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 16528A-001810  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 384:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-464-538B-384

Query Match 29.6%; Score 42; DB 2; Length 17;  
Best Local Similarity 63.6%; Pred. No. 7.5;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 6 TAWYQHDRLAY 16  
DB 7 TAWYDHFLLSY 17

Search completed: August 15, 2002, 11:59:43

Job time: 248 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 12:00:14 ; Search time 25.35 seconds  
(without alignments)  
90.972 Million cell updates/sec

Title: DEVI-613-COMBIFRAG  
Perfect score: 142  
Sequence: 1 SRVPTAWYQHDLRATRYRFRVRRP 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	39.4	133	2 C31211	T-cell receptor be
2	56	39.4	134	2 A49035	T-cell receptor be
3	56	39.4	135	2 B28823	T-cell receptor be
4	56	39.4	136	1 RWMSE1	T-cell receptor be
5	56	39.4	137	2 I71937	MHC class II I-A-b
6	54	38.0	345	2 T35357	hypothetical prote
7	52.5	37.0	267	2 S52347	hypothetical prote
8	52	36.6	85	2 JH0346	T-cell receptor be
9	51	35.9	389	2 H70841	probable amha prot
10	51	35.9	454	2 T26654	hypothetical prote
11	51	35.9	833	2 F81989	hypothetical prote
12	49.5	34.9	170	2 A12961	conserved hypotet
13	49.5	34.9	175	2 E98321	hypothetical prote
14	49.5	34.5	417	2 T33376	hypothetical prote
15	48.5	34.2	434	2 B69271	hypothetical prote
16	48.5	34.2	506	2 A81191	hypothetical prote
17	48	33.8	269	2 A95148	llycd protein (imp
18	48	33.8	269	2 G98015	llycd protein (imp
19	48	33.8	295	2 F70936	hypothetical prote
20	47	33.1	200	2 T34654	probable transmem
21	46.5	32.7	33	2 S28842	madulin II chain
22	46.5	32.7	135	2 JC5379	madulin II chain
23	46.5	32.7	1186	2 T03180	tyrosine protein k
24	46	32.4	461	2 A10279	hypothetical prote
25	45.5	32.0	382	2 E83322	probable PAD-depen
26	45.5	32.0	2290	1 GNNYE	genome polyprotein
27	45.5	32.0	2292	2 S35961	capsid polyprotein
28	45	31.7	253	2 A75252	conserved hypotet
29	45	31.7	302	2 AH2196	hypothetical prote

30	45	31.7	416	2 D83386	hypothetical prote
31	45	31.7	470	2 S47892	neutral amino acid
32	45	31.7	663	2 H82731	glutaryl-7-ACA acy
33	45	31.7	790	2 T34293	hypothetical prote
34	45	31.7	1776	2 G86280	protein T5E21.13 (
35	44.5	31.3	405	2 S61551	breast-regressing
36	44	31.0	62	2 S72792	hypothetical prote
37	44	31.0	125	2 F71276	holo-facil-carrier
38	44	31.0	209	2 A39759	photosystem I 18k
39	44	31.0	216	2 S35151	photosystem I cha
40	44	31.0	244	2 S50685	hypothetical prote
41	44	31.0	309	2 F83605	probable permease
42	44	31.0	314	2 B98232	probable 2-hydroxy
43	44	31.0	327	2 AC3054	2-hydroxyacid deny
44	44	31.0	354	2 S15660	(2'-5')oligoc(A) sy
45	44	31.0	476	2 T43464	hypothetical prote

## ALIGNMENTS

RESULT 1  
C31211  
T-cell receptor beta chain precursor V region (V2-D1.1-J1.2) - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 23-Jul-1999  
C:Accession: C31211  
R:Malissen, M.; Trucy, J.; Letourneur, F.; Rebbi, N.; Dunn, D.E.; Fitch, F.W.; Hood, Cell 55, 49-59, 1988  
A>Title: A T cell clone expresses two T cell receptor alpha genes but uses one alpha-  
A:Reference number: A94656; MUID:89003051  
A:Accession: C31211  
A:Molecule type: DNA  
A:Residues: 1-133 <MAL>  
A:Cross-references: GB:M22605; NID:q340718; PIDN:AAA63778.1; PID:q710557  
C:Genetics:  
A:Introns: 19/1  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: T-cell receptor

Query Match 39.4%; Score 56; DB 2; Length 133;  
Best Local Similarity 57.1%; Pred. No. 0.9;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 SRVPTAWYQHDLR 14  
DB 49 SQYPTAWYQDDIQ 62

RESULT 2  
A49035  
T-cell receptor beta chain variable region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Dec-1993 #sequence\_revision 25-Apr-1997 #text\_change 23-Jul-1999  
C:Accession: A49035  
R:Glebas, A.; Hansen-Hagge, T.; Von Bonin, A.; Wetzien, H.U. Eur. J. Immunol. 22, 335-341, 1992  
A>Title: Increased frequency of 2,4,6-trinitrophenyl (TNP)-specific, H-2b-restricted trieted, TNP-specific cytolytic T cell clone.  
A:Reference number: A49035; MUID:92164708  
A:Accession: A49035  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-134 <TGL>  
A:Cross-references: GB:S85477; NID:q246280; PIDN:AAB21544.1; PID:q246281  
A>Note: sequence extracted from NCBI backbone (NCBI:85477, NCBI:P:85529)  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: T-cell receptor

Query Match 39.4%; Score 56; DB 2; Length 134;  
Best Local Similarity 57.1%; Pred. No. 0.91;

```
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 SRPWTANYQHDLR 14
   |::||| |
Db 49 SQYPWMSWYQODLQ 62

RESULT 3
B28823
T-cell receptor beta chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 23-Jul-1999
C:Accession: B28823
C:Title: K.N.; Dattlof, B.M.; Gilmore, J.A.; Kronman, A.C.; Lee, J.H.; Maxam, A.M.; Rao, A.
Cell 54, 247-261, 1988
A:Title: The T cell receptor V-alpha-3 gene segment is associated with reactivity to p-g
A:Reference number: A90900; MUID:88270504
A:Accession: B28823
A:Molecule type: DNA
A:Residues: 1-135 <TRAN>
A:Cross-references: GB:M21203; NID:g201383; PIDN:AAA40254.1; PID:g554307
C:Genetics:
A:Introns: 18/3
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: T-cell receptor

Query Match 39.4%; Score 56; DB 2; Length 135;
Best Local Similarity 57.1%; Pred. No. 0.91;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 SRPWTANYQHDLR 14
   |::||| |
Db 49 SQYPWMSWYQODLQ 62

RESULT 4
RWMSE1
T-cell receptor beta chain precursor V region (E1) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 22-Jun-1999
C:Accession: A02009
R:Patten, P.; Yokota, T.; Rothbard, J.; Chien, Y.; Arai, K.; Davis, M.M.
Nature 312, 40-46, 1984
A:Title: Structure, expression and divergence of T-cell receptor beta-chain variable reg
A:Reference number: A93346; MUID:85036636
A:Accession: A02009
A:Molecule type: mRNA
A:Residues: 1-136 <PAT>
A:Cross-references: GB:X01642; NID:g54308; PIDN:CAA25799.1; PID:g762989
C:Comment: This sequence was derived from a T-helper clone.
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; receptor; T-cell
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-136/Product: T-cell receptor beta chain precursor V region (E1) #status predicted <
F:21-116/Region: V segment
F:117-120/Region: D segment
F:121-136/Region: J segment
F:44-112/Disulfide bonds: #status predicted

Query Match 39.4%; Score 56; DB 1; Length 136;
Best Local Similarity 57.1%; Pred. No. 0.92;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 SRPWTANYQHDLR 14
   |::||| |
Db 49 SQYPWMSWYQODLQ 62

RESULT 5
I71937
MHC class II I-A-beta protein precursor - mouse (fragment)
```

```
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 23-Jul-1999
C:Accession: I71937
R:Spinelia, D.G.; Hansen, T.H.; Walsh, W.D.; Behlke, M.A.; Tillinghast, J.P.; Chou, H
J. Immunol. 138, 3991-3995, 1987
A:Title: Receptor diversity of insulin-specific T cell lines from C57BL (H-2b) mice.
A:Reference number: I55978; MUID:87224052
A:Accession: I71937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-137 <RES>
A:Cross-references: GB:M16680; NID:g199471; PIDN:AAA39625.1; PID:g199472
C:Superfamily: immunoglobulin V region; immunoglobulin homology

Query Match 39.4%; Score 56; DB 2; Length 137;
Best Local Similarity 57.1%; Pred. No. 0.92;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 SRPWTANYQHDLR 14
   |::||| |
Db 49 SQYPWMSWYQODLQ 62

RESULT 6
I73357
hypothetical protein SC5H1.41 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T73357
R:Olliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z21575
A:Accession: T73357
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-345 <OLIT>
A:Cross-references: EMBL:AL049863; PIDN:CAB42966.1; GSPDB:GN00070; SCOEDB:SC5H1.41
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5H1.41

Query Match 38.0%; Score 54; DB 2; Length 345;
Best Local Similarity 45.0%; Pred. No. 4.3;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 5 WTAWYQHDLRARFVRRP 24
   |::||| |
Db 82 WAERSRHEIGAYRSPVRRP 101

RESULT 7
S52347
hypothetical protein 1 - Lactobacillus leichmannii
C:Species: Lactobacillus leichmannii
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 05-Nov-1999
C:Accession: S52347
R:Schenk-Groeninger, R.
submitted to the EMBL Data Library, January 1995
A:Reference number: S52347
A:Accession: S52347
A:Molecule type: DNA
A:Residues: 1-267 <SCH>
A:Cross-references: EMBL:X81869; NID:g666067; PIDN:CAA57458.1; PID:g666068
A:Experimental source: DSM 20076

Query Match 37.0%; Score 52.5; DB 2; Length 267;
Best Local Similarity 52.4%; Pred. No. 5.4;
Matches 11; Conservative 2; Mismatches 5; Indels 3; Gaps 1;
QY 2 RVPWTANYQHDLRAY--RRF 19
```

Db 161 QVPLAVYVNHDLRYNCCRRF 181

RESULT 8

JH0346  
T-cell receptor beta chain V region (MT1-6) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 23-Jul-1999  
C:Accession: JH0346  
R:Taylor, A.H.; Haberman, A.M.; Gerhard, W.; Caton, A.J.  
J. Exp. Med. 172, 1643-1651, 1990  
A:Title: Structure-function relationships among highly diverse T cells that recognize a  
A:Reference number: JH0333; M0ID:91079767  
A:Accession: JH0346  
A:Molecule type: mRNA  
A:Residues: 1-85 <TAY>  
A:Cross-references: GB:M34203; NID:g201A18; PIDN:AAA58763.1; PID:g201A19  
C:Comment: This T-cell receptor recognizes a determinant from Influenza virus hemagglutinin  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: receptor; T-cell

Query Match 36.6%; Score 52; DB 2; Length 85;  
Best Local Similarity 63.0%; Pred. No. 2.1;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 4 PWTAWYQHDLRYRRFVRRP 14  
| | : | | | : |  
Db 3 PMSWYQDDLD 13

RESULT 9

H70841  
probable amIA protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70841  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Ruter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; M0ID:98295987  
A:Accession: H70841  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-389 <COL>  
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAAI7077.1; PID:g289421  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: amIA  
C:Superfamily: hippurate hydrolase

Query Match 35.9%; Score 51; DB 2; Length 389;  
Best Local Similarity 45.0%; Pred. No. 13;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Oy 5 WTAWYQHDLRYRRFVRRP 24  
| | : | | | : |  
Db 10 WLAHNDLWGRHRRHRRP 29

RESULT 10

T26554  
hypothetical protein Y38E10A.d - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26554  
R:Wallis, J.  
submitted to the EMBL Data Library, September 1999

A:Reference number: Z20252  
A:Accession: T26654  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-454 <WTL>  
A:Cross-references: EMBL:AL110484; NID:el542205; PIDN:CAB54395.1; CESP:Y38E10A.d  
C:Genetics:  
A:Experimental source: clone Y38E10A  
A:Gene: CESP:Y38E10A.d  
A:Introns: 67/3; 115/1; 159/1; 391/3; 434/2

Query Match 35.9%; Score 51; DB 2; Length 454;  
Best Local Similarity 41.7%; Pred. No. 15;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Oy 1 SRVPTAWYQHDLRYRRFVRRP 24  
| | : | | | : |  
Db 258 SWPTGSWYVPCAKNSFLCKRP 281

RESULT 11

F81989  
hypothetical protein NMA0692 [imported] - Neisseria meningitidis (strain Z2491 serogr  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: F81989  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandrea  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: AB1775; M0ID:20222556  
A:Accession: F81989  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-833 <PAR>  
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83978.1; PID:g737  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0692

Query Match 35.9%; Score 51; DB 2; Length 833;  
Best Local Similarity 33.3%; Pred. No. 26;  
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Oy 2 RVPTAWYQHDLRYRRFVRRP 22  
| | : | | | : |  
Db 666 RTGEPTWRHDVYRQYRNO 686

RESULT 12

AI2961  
conserved hypothetical protein Atu3295 [imported] - Agrobacterium tumefaciens (strain  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AI2961  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guentlner, D.; Kutayavin, T.; Levy, R.; Li, M.; MCL  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AI2961  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-170 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AML4411.1; PID:g17741680; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu3295

A:Map position: linear chromosome

Query Match 34.9%; Score 49.5; DB 2; Length 170;  
Best Local Similarity 50.0%; Pred. No. 9.3;  
Matches 10; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 4 PWT-AMYQHDLRAYRRFVHR 22  
DB 134 PWTNAMYDYCSQRYRSENSR 153

RESULT 13

hypothetical protein AGR\_L\_3047 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: E98321  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: E98321  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-175 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK90095.1; PID:q1516082; GSPDB:GN00170  
A:Gene: AGR\_L\_3047  
A:Map position: linear chromosome

Query Match 34.9%; Score 49.5; DB 2; Length 175;  
Best Local Similarity 50.0%; Pred. No. 9.5;  
Matches 10; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 4 PWT-AMYQHDLRAYRRFVHR 22  
DB 139 PWTNAMYDYCSQRYRSENSR 158

RESULT 14

hypothetical protein T07H3.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T33376  
R:Beck, C.; O'Brien, D.; Kramer, J.  
Submitted to the EMBL Data Library, July 1998  
A:Description: The sequence of C. elegans cosmid T07H3.  
A:Reference number: 221332  
A:Accession: T33376  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-417 <BEC>  
A:Cross-references: EMBL:AF077540; PIDN:AAC26308.1; GSPDB:GN00020; CESP:T07H3.5  
A:Experimental source: strain Bristol N2; clone T07H3  
C:Genetics:  
A:Gene: CESP:T07H3.5  
A:Map position: 2  
A:introns: 65/3; 152/1

Query Match 34.5%; Score 49; DB 2; Length 417;  
Best Local Similarity 39.1%; Pred. No. 26;  
Matches 9; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 2 RVPWTAWYQHD-LRAYRRFVHR 24  
DB 256 RQPWGLWYTGNCYDRKRFCKRP 278

RESULT 15

hypothetical protein AF0170 - Archaeoglobus fulgidus  
B69271  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: B69271  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
.,; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
Glodek, A.; Zhou, L.; Overbeek, R.; Cocayne, J.D.; Meldman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes,  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
A:Reference number: B69271  
A:Accession: B69271  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-434 <KLE>  
A:Cross-references: GB:AE001094; GB:AE000782; NID:g2689417; PIDN:AAB91065.1; PID:g265

Query Match 34.2%; Score 48.5; DB 2; Length 434;  
Best Local Similarity 42.9%; Pred. No. 31;  
Matches 9; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

QY 3 VPWTAWYQHD-LRAYRRFVHR 22  
DB 162 IPNVLMFKHDEIRALRYLRLR 182

Search completed: August 15, 2002, 12:00:15  
Job time: 205 sec



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 12:05:55 ; Search time 16.31 Seconds  
(without alignments)  
56.975 Million cell updates/sec

Title: DEVI-613-COMBIFRAG  
Perfect score: 142  
Sequence: 1 SRVPTAWYQHDRLRAYRRFVHRP 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	56	39.4	136	TVB6_MOUSE
2	46.5	32.7	155	2SS2_CAPMA
3	45.5	32.0	2290	POLG_EMCV
4	45	31.7	470	MTR_NEUCR
5	44	31.0	125	ACPS_TREPA
6	44	31.0	209	PSAL_HORVU
7	44	31.0	216	PSAL_SPTOL
8	44	31.0	244	YE12_YEAST
9	44	31.0	586	ULB4_HCMVA
10	44	31.0	587	ULB4_HCMVT
11	44	31.0	669	ACSA_PENCH
12	44	31.0	775	MCDL_MOUSE
13	44	31.0	775	MCDL_MOUSE
14	44	31.0	778	MCDL_MOUSE
15	44	31.0	2485	MGDI_HUMAN
16	44	30.3	90	PTND_HUMAN
17	43	30.3	217	YPTJ_ECOLI
18	43	30.3	310	PSAL_CUCSA
19	43	30.3	367	DRNG_RAT
20	43	30.3	367	OASA_MOUSE
21	43	30.3	436	CP22_HORVU
22	43	30.3	515	AMYP_RAT
23	43	30.3	515	YJEF_ECOLI
24	43	30.3	1108	EMBA_MYCAV
25	43	30.3	1139	VRNA_BSMV
26	43	30.3	1257	PCGN_RAT
27	43	30.3	1268	PCGN_MOUSE
28	42.5	29.9	378	ACE_MOUSE
29	42.5	29.9	624	071A_DROME
30	42.5	29.9	624	HTPG_ECOLI
31	42.5	29.9	624	HTPG_SALTI
32	42	29.6	305	HTPG_SALTY
33	42	29.6	398	DRNG_HUMAN
34	42	29.6	398	KR2_PPVN3

34	42	29.6	677	1	BGAL_HUMAN	P16278	homo sapien
35	42	29.6	770	1	STAG_HUMAN	P40763	homo sapien
36	42	29.6	770	1	STAG_MOUSE	P42237	mus musculus
37	42	29.6	770	1	STAG_RAT	P52631	rattus norv
38	42	29.6	1310	1	ACE_RABIT	P12822	oryctolagus
39	41.5	29.2	721	1	GLGX_MYCTU	Q10767	mycobacteri
40	41.5	29.2	1219	1	YMAZ_YEAST	Q04439	saccharomyc
41	41	28.9	107	1	HMAD_AMBME	P50210	ambystoma m
42	41	28.9	127	1	LYC_COLLI	P00708	columba liv
43	41	28.9	129	1	LYC2_MYCPN	P75406	mycoplasma
44	41	28.9	144	1	LYC2_ONCMY	P11941	oncorhynch
45	41	28.9	158	1	RACR_ECOLI	P76062	escherichia

## ALIGNMENTS

RESULT	ID	TVB6_MOUSE	STANDARD	PRT	136 AA.
AC	P04214	TVB6_MOUSE	STANDARD	PRT	136 AA.
DT	20-MAR-1987	(Rel. 04, Created)			
DT	20-MAR-1987	(Rel. 04, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	T-cell receptor beta chain V region E1 precursor.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=8503636; PubMed=6092964;				
RA	Patten P., Yokota T., Rothbard J., Chien Y., Aral K., Davis M.M.;				
RT	"Structure, expression and divergence of T-cell receptor beta chain variable regions."				
RL	Nature 312:40-46(1984).				
CC	-I- MISCELLANEOUS: THIS SEQUENCE WAS DERIVED FROM A T-HELPER CLONE.				
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	CC EMBL, X01642; CA25799.1; -				
DR	PIR: A02009; RMMSE1.				
DR	InterPro: IPR003599; Ig.				
DR	InterPro: IPR003006; Ig_MHC.				
DR	Pfam: PF000047; Ig_1.				
DR	SMART: SM00409; Ig_1.				
DR	T-cell; Receptor; Signal.				
KW	SIGNAL				
FT	CHAIN	21	136		T-CELL RECEPTOR BETA CHAIN V REGION E1.
FT	DOMAIN	21	116		V SEGMENT.
FT	DOMAIN	117	120		D SEGMENT.
FT	DOMAIN	121	136		J SEGMENT.
FT	DISULFID	44	112		BY SIMILARITY.
FT	NON_TER	136	136		
SO	SEQUENCE	136 AA;	15538 MW;	2B93C4912E1B705 CRC64;	

Query Match 39.4%; Score 56; DB 1; Length 136;  
Best Local Similarity 57.1%; Pred. NO. 0.14;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 1 SRVPTAWYQHDRL 14  
1: ||| ||| |||  
Db 49 SQYPMWYQDDLO 62

RESULT 2

2SS2\_CAPMA STANDARD: PRT: 155 AA.  
 ID 2SS2\_CAPMA 004774;  
 AC P30233; 004774;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mabinlin II precursor (MAB II) (Sweet protein).  
 OS Capparis masaikei (Mabinlin).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Capparis.  
 OC NCBI\_TaxId=13395;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Seed;  
 RA MEDLINE=97128796; PubMed=8973336;  
 RA Nirasawa S., Masuda Y., Nakaya K., Kurihara Y.;  
 RT "Cloning and sequencing of a cDNA encoding a heat-stable sweet  
 RT protein, mabinlin II.";  
 RL gene 181:225-227(1996).  
 RN [2]  
 RP SEQUENCE OF 36-68 AND 83-154.  
 RC TISSUE=Seed;  
 RA MEDLINE=93145958; PubMed=8425538;  
 RA Liu X., Maeda S., Hu Z., Aiuchi T., Nakaya K., Kurihara Y.;  
 RT "Purification, complete amino acid sequence and structural  
 RT characterization of the heat-stable sweet protein, mabinlin II.";  
 RL Eur. J. Biochem. 211:281-287(1993).  
 RN [3]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RA MEDLINE=94002261; PubMed=8399391;  
 RA Nirasawa S., Liu X., Nishino T., Kurihara Y.;  
 RT "Disulfide bridge structure of the heat-stable sweet protein mabinlin  
 RT II.";  
 RL Biochim. Biophys. Acta 1202:277-280(1993).  
 CC -I- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-  
 CC INDUCING ACTIVITY.  
 CC -I- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY  
 CC DISULFIDE BONDS.  
 CC -I- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D83997; BAA12204.1;  
 DR Mendel: 14634; Capma.1175.14634.  
 DR InterPro: IPR003612; AAI.  
 DR InterPro: IPR000617; Napin.  
 DR Pfam: PF01631; Seedstore\_2S; 1.  
 DR PRINTS: PR00496; Napin.  
 DR PRODOM: P0002498; Napin; 1.  
 DR SMART: SM00499; AAI; 1.  
 KM Seed storage protein; Albumin; Signal; Sweet-taste.  
 FT SIGNAL 1 20  
 FT PROPEP 21 35  
 FT CHAIN 36 68 MABINLIN II, A CHAIN.  
 FT PROPEP 69 82  
 FT CHAIN 83 154 MABINLIN II, B CHAIN.  
 FT PROPEP 155 155  
 FT DISULFID 40 103  
 FT DISULFID 53 92  
 FT DISULFID 93 141  
 FT DISULFID 105 149  
 FT MOD\_RES 36 83 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 83 86 PYRROLIDONE CARBOXYLIC ACID.  
 FT CONFLICT 148 148 A -> T (IN REF. 1).  
 FT CONFLICT 153 153 A -> T (IN REF. 1).  
 SQ SEQUENCE 155 AA: 18089 MW: 725855DECC2D46A CRC64;

Query Match 32.7%; Score 46.5; DB 1; Length 155;  
 Best Local Similarity 66.7%; Pred. No. 4.2;  
 Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
 QY 10 QND-LRAYRFPVHRR 23  
 DB 46 QNDLRACQRFTHRR 60  
 RESULT 3  
 POLG\_EMCV STANDARD: PRT: 2290 AA.  
 ID POLG\_EMCV  
 AC P03304;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polyprotein (Contains: Coat proteins VP1 TO VP4; Core proteins  
 DE P2A TO P2C; P3A; Genome-linked protein VP6; Picornain 3C  
 DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D  
 DE (EC 2.7.7.48)).  
 OS Encephalomyocarditis virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Cardiovirus.  
 OC NCBI\_TaxId=12104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=84169586; PubMed=6324136;  
 RA Palmenberg A.C., Kirby E.M., Janda M.R., Drake N.L., Duke G.M.,  
 RA Potratz K.F., Collett M.S.;  
 RT "The nucleotide and deduced amino acid sequences of the  
 RT encephalomyocarditis viral polyprotein coding region.";  
 RL Nucleic Acids Res. 12:2969-2985(1984).  
 CC -I- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN  
 CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.  
 CC -I- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
 CC VP3, AND VP4.  
 CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X00463; CAA25152.1; -.  
 DR PIR: A03906; GNYE.  
 DR HSSP: P12296; 2MEV.  
 DR MEROPS: C03.009; -.  
 DR MEROPS: U29.001; -.  
 DR InterPro: IPR000605; RNA\_helicase.  
 DR InterPro: IPR001205; RNA\_pol\_P3D.  
 DR InterPro: IPR001676; Rny.  
 DR Pfam: PF00073; Rny; 3.  
 DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam: PF00910; RNA\_helicase; 1.  
 KM RNA-directed RNA polymerase; Hydrolyase; Thiol protease; Myristate.  
 FT PROPEP 1 67  
 FT CHAIN 68 136 COAT PROTEIN VP4 (RHQ).  
 FT CHAIN 137 391 COAT PROTEIN VP2 (BETA).  
 FT CHAIN 392 622 COAT PROTEIN VP3 (GAMMA).  
 FT CHAIN 623 910 COAT PROTEIN VP1 (ALPHA).  
 FT CHAIN 911 1056 CORE PROTEIN P2A (G).  
 FT CHAIN 1057 1192 CORE PROTEIN P2B (I).  
 FT CHAIN 1193 1517 CORE PROTEIN P2C (F).  
 FT CHAIN 1518 1605 CORE PROTEIN P3A.  
 FT CHAIN 1606 1625 GENOME-LINKED PROTEIN VP6 (H).

FT CHAIN 1626 1830 PICORNAIN 3C (P22).  
 FT CHAIN 1831 2290 RNA-DIRECTED RNA POLYMERASE P3D (E).  
 FT LIPID 68 68 MYRISTATE (BY SIMILARITY).  
 FT ACT\_SITE 1784 1784 PROTEASE (POTENTIAL).  
 FT ACT\_SITE 1802 1802 PROTEASE (POTENTIAL).  
 SQ SEQUENCE 2290 AA; 255756 MW; 26BC81B7CF68C5 CRC64;

Query Match 32.0%; Score 45.5; DB 1; Length 2290;  
 Best Local Similarity 55.6%; Pred. No. 99;  
 Matches 10; Conservative 0; Mismatches 5; Indels 3; Gaps 1;

QY 1 SRVPTAM---YOHDLRA 15  
 |||||  
 Db 964 SRAPWNPKNYQAVLRA 981

RESULT 4  
 MTR\_NEUCR STANDARD; PRT; 470 AA.  
 AC P38680;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE N amino acid transport system protein (Methyltryptophan resistance protein).  
 GN MTR.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_Taxid=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OAK RIDGE;  
 RX MEDLINE=95095055; PubMed=1838345;  
 RA Dillon D., Stadler D.;  
 RT "Spontaneous mutation at the mtr locus in neurospora: the molecular spectrum in wild-type and a mutator strain."  
 RL Genetics 138:61-74(1994).  
 RN [2]  
 RP SEQUENCE OF 210-470 FROM N.A.  
 RX MEDLINE=92146948; PubMed=1838345;  
 RA Koo K., Stuart W.D.;  
 RT "Sequence and structure of mtr, an amino acid transport gene of Neurospora crassa."  
 RL Genome 34:644-651(1991).  
 CC -1- FUNCTION: REQUIRED FOR THE TRANSPORT OF NEUTRAL ALIPHATIC AND ANOMATIC AMINO ACIDS VIA THE N SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L34605; AAA33600.1; -;  
 DR EMBL: S81767; AAB21410.1; -;  
 DR PIR: A54551; A54551.  
 DR InterPro: IPR002422; AA\_rel\_permease\_2.  
 DR Pfam: PF01490; Aa\_trans\_1.  
 KW Transport; Amino-acid transport; Transmembrane.  
 FT DOMAIN 1 56 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 57 77 POTENTIAL.  
 FT TRANSMEM 78 98 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 99 131 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 132 152 POTENTIAL.  
 FT DOMAIN 153 168 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 169 189 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT DOMAIN 212 236 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 237 257 POTENTIAL.  
 FT DOMAIN 258 275 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 276 296 POTENTIAL.  
 FT DOMAIN 297 316 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 317 337 POTENTIAL.  
 FT DOMAIN 338 357 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 358 378 POTENTIAL.  
 FT DOMAIN 379 386 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 387 407 POTENTIAL.  
 FT DOMAIN 408 427 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 428 448 POTENTIAL.  
 FT DOMAIN 449 470 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 470 AA; 51162 MW; E8132D1A62373300 CRC64;

Query Match 31.7%; Score 45; DB 1; Length 470;  
 Best Local Similarity 38.9%; Pred. No. 22;  
 Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 VPWTAWYOHDLRAVRRREV 20  
 |||||  
 Db 227 VPWSCPKEDLSLAEFT 244

RESULT 5  
 ACPS\_TREPA STANDARD; PRT; 125 AA.  
 ID ACPS\_TREPA  
 AC O83800;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase).  
 GN ACPS OR TP0828.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_Taxid=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOLS;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G., Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J., Khalak H., Richardson D., Howell J.K., Chidambaram M., Utersback T., McDonald L., Artlisch P., Bowman C., Cotton M.D., Fujii C., Garland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C.;  
 RA "Complete genome sequence of Treponema pallidum, the syphilis spirochete."  
 RT Science 281:375-388(1998).  
 CC -1- FUNCTION: TRANSFERS THE 4'-PHOSPHOPANTHETINE MOIETY FROM COENZYME A TO A SER OF ACYL-CARRIER PROTEIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine 3',5'-bisphosphate + holo-[acyl-carrier protein].  
 CC -1- SIMILARITY: BELONGS TO THE ACPS FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE001253; AAC65794.1; -;  
 DR TIGR: TP0828; -;  
 DR InterPro: IPR002582; ACPS.  
 DR Pfam: PF01648; ACPS; 1.  
 DR ProDom: PD004282; ACPS; 1.  
 KW Lipid synthesis; Transferease; Complete proteome.  
 SQ SEQUENCE 125 AA; 14118 MW; 88A05B5B0D20821F CRC64;

Query Match 31.0%; Score 44; DB 1; Length 125;  
 Best Local Similarity 45.0%; Pred. No. 7.9;  
 Matches 9; Conservative 3; Mismatches 4; Indels 4; Gaps 1;  
 OY 3 PWTAWYQHDLRAYRRFVR 22  
 1 11 1:1 111 1:  
 DB 16 VSWT-----HNVRLRRFFHQ 31

RESULT 6  
 ID PSAL\_HORVU STANDARD; PRT: 209 AA.  
 AC P23933;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Photosystem I reaction center subunit XI, chloroplast precursor  
 DE (PSI-L) (PSI subunit V).  
 GN PSAL.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Hordeum.  
 OC NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 51-70 AND 72-74.  
 RC STRAIN=CV. SVALOF'S BONUS;  
 RA MEDLINE=91201317; PubMed-2016292;  
 RA Okels J.S., Scheeller H.V., Svendsen I., Moeller B.L.;  
 RT "Isolation and characterization of a cDNA clone encoding an 18-kDa  
 RT hydrophobic photosystem I subunit (PSI-L) from barley (Hordeum  
 RT vulgare L.).", 266:6767-6773(1991).  
 RL J. Biol. Chem. 266:6767-6773(1991).  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast  
 CC thylakoid membrane (Probable).  
 CC -1 INDUCTION: BY LIGHT.  
 CC -1 SIMILARITY: BELONGS TO THE PSAL FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M61146; AAA62700.1; -;  
 DR PIR: A39759; A39759.  
 DR InterPro: IPR003757; Psal.  
 DR Pfam: PF02605; Psal; 1.  
 KW Photosystem I; Photosynthesis; Chloroplast; Transit peptide;  
 KM Thylakoid; Transmembrane; Acetylation.  
 FT TRANSIT 1 40  
 FT CHAIN 41 209  
 FT DOMAIN 41 124 PHOTOSYSTEM I REACTION CENTER SUBUNIT XI.  
 FT TRANSMEM 125 146 POTENTIAL.  
 FT DOMAIN 147 179 LUMENAL (POTENTIAL).  
 FT TRANSMEM 180 202 POTENTIAL.  
 FT DOMAIN 203 209 STROMAL (POTENTIAL).  
 FT MOD\_RES 41 41 ACETYLATION (PROBABLE).  
 FT SEQUENCE 209 AA; 22211 MW; 95741FD3F290F9E1 CRC64;

Query Match 31.0%; Score 44; DB 1; Length 209;  
 Best Local Similarity 52.9%; Pred. No. 13;  
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 4 PWTAWYQHDLRAYRRFV 20  
 1 111 1:1 111 1:  
 DB 72 PLVAVYLSMLPAYRTAV 88

RESULT 7

PSAL\_SPTOL STANDARD; PRT: 216 AA.  
 ID PSAL\_SPTOL Q41385;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Photosystem I reaction center subunit XI, chloroplast precursor  
 DE (PSI-L) (PSI subunit V).  
 GN PSAL.  
 OS Spinacia oleracea (Spinach).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
 OC NCBI\_TaxID=3562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. MONATOL;  
 RA MEDLINE=93344519; PubMed=8343606;  
 RA Flieger K., Oelmueller R., Herrmann R.G.;  
 RT "Isolation and characterization of cDNA clones encoding a 18.8 kDa  
 RT polypeptide, the product of the gene psal, associated with  
 RT photosystem I reaction center from spinach.",  
 RL Plant Mol. Biol. 22:703-709(1993).  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast  
 CC thylakoid membrane (Probable).  
 CC -1 SIMILARITY: BELONGS TO THE PSAL FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X64445; CAA5775.1; -;  
 DR InterPro: IPR003757; Psal.  
 DR Pfam: PF02605; Psal; 1.  
 KW Photosystem I; Photosynthesis; Chloroplast; Transit peptide;  
 KM Thylakoid; Transmembrane.  
 FT TRANSIT 1 47  
 FT CHAIN 48 216  
 FT DOMAIN 48 134 STROMAL (POTENTIAL).  
 FT TRANSMEM 135 155 POTENTIAL.  
 FT DOMAIN 156 188 LUMENAL (POTENTIAL).  
 FT TRANSMEM 189 209 POTENTIAL.  
 FT DOMAIN 210 216 STROMAL (POTENTIAL).  
 FT SEQUENCE 216 AA; 22937 MW; 603DCA983C7C383B CRC64;

Query Match 31.0%; Score 44; DB 1; Length 216;  
 Best Local Similarity 50.0%; Pred. No. 14;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 4 PWTAWYQHDLRAYRRFV 21  
 1 111 1:1 111 1:  
 DB 79 PLVAVYLSMLPAYRTAVN 96

RESULT 8  
 ID YE12\_YEAST STANDARD; PRT: 244 AA.  
 AC P40098;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical 27.7 kDa protein in ISC10 3' region.  
 GN YER182W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OC NCBI\_TaxID=4932;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Dierich F.S., Mulligan J.T., Hennessey R.M., Allen E., Araujo R.,
RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBD databases.
CC -1- SIMILARITY: SOME, TO S.POMBE SPAC3A12.08.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U18922; AAB64709.1; -
DR SCD: S0000984; YER182W.
KM Hypothetical protein; Transmembrane.
FT TRANSMEM 29 49
FT TRANSMEM 139 159 POTENTIAL.
SQ SEQUENCE 244 AA; 27698 MW; A78338C86EA52766 CRC64;

Query Match
Best Local Similarity 31.0%; Score 44; DB 1; Length 244;
Matches 12; Conservative 2; Mismatches 3; Indels 14; Gaps 3;

OY 3 VPWT-----AWY--OH---DLRAYRRF 19
Db 30 IPWTIRGSGFLGSGWYLDLQHHMTFTDLAYRRY 60

RESULT 9
UL84_HCMVA STANDARD; PRT; 586 AA.
AC P16727;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE 65 kDa early nonstructural protein (UL84 protein).
GN UL84.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Predile E., Satchell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X17403; CAA35358.1; -
DR PIR: S09848; WMBDE.
KM Nonstructural protein.
FT DOMAIN 9 19
FT DOMAIN 162 170
FT DOMAIN 171 183
FT DOMAIN 162 170 LYS-RICH (BASIC).
FT DOMAIN 170 183 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 162 AA; 170 MW; LYS-RICH (BASIC).

```

```

FT DOMAIN 171 182 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 586 AA; 65428 MW; 54AB912D6077223F CRC64;

Query Match
Best Local Similarity 31.0%; Score 44; DB 1; Length 586;
Matches 11; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

OY 2 RVPW--TAWQHLRAYRRF 19
Db 427 RVPVELTKNSHRLRYRRF 446

RESULT 10
UL84_HCMVT STANDARD; PRT; 587 AA.
AC P29839;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE 65 kDa early nonstructural protein (UL84 protein).
GN UL84.
OS Human cytomegalovirus (strain Towne).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114132; PubMed=1309892;
RA He Y.S., Xu L., Huang E.S.;
RT "Characterization of human cytomegalovirus UL84 early gene and
RT identification of its putative protein product."
RL J. Virol. 66:1098-1108(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M81432; AAA45947.1; -
DR PIR: A41808; WMBETE.
KM Nonstructural protein.
FT DOMAIN 9 19
FT DOMAIN 162 170 LYS-RICH (BASIC).
FT DOMAIN 171 183 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 587 AA; 65388 MW; 13C170E41FB3220B CRC64;

Query Match
Best Local Similarity 31.0%; Score 44; DB 1; Length 587;
Matches 11; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

OY 2 RVPW--TAWQHLRAYRRF 19
Db 428 RVPVELTKNSHRLRYRRF 447

RESULT 11
ACSA_PENCH STANDARD; PRT; 669 AA.
AC P36333;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A synthetase (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-
DE activating enzyme).
GN PACA OR ACUA.
OS Penicillium chrysogenum.
OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiaceae; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

```

```

OX  NCBI_TaxID=5076;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93366184; PubMed=8103029;
RA  Matinez-Blanco H., Orejas M., Reglero A., Luengo J.M., Penalva M.A.;
RT  "Characterisation of the gene encoding acetyl-CoA synthetase in
RT  penicillium chrysogenum: conservation of intron position in
RT  plectomyces.";
RL  Gene 130:265-270(1993).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93159753; PubMed=7765289;
RA  Gouka R.J., van Hartingsveldt W., Bovenberg R.A., van Zeijl C.M.,
RA  van den Hondel C.A., van Gorcom R.F.;
RT  "Development of a new transformant selection system for Penicillium
RT  chrysogenum: isolation and characterization of the P. chrysogenum
RT  acetyl-coenzyme A synthetase gene (faca) and its use as a homologous
RT  selection marker.";
RL  Appl. Microbiol. Biotechnol. 38:514-519(1993).
CC  -I- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
CC  acetyl-CoA.
CC  -I- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC  FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; S54801; AAC60546.1; -
DR  EMBL; L09598; AAA02921.1; -
DR  PIR; JN0781; JN0781.
DR  HSP; P08659; JICI.
DR  InterPro; IPR00873; AMP-bind.
DR  Pfam; PF00501; AMP-binding; 1.
DR  PRINTS; PR00154; AMPBINDING.
DR  PROSITE; PS00455; AMP_BINDING; 1.
KW  Ligase.
SO  SEQUENCE 669 AA; 74287 MW; 7C05220D321D736C CRC64;

Query Match      31.0%; Score 44; DB 1; Length 669;
Best Local Similarity 29.2%; Pred. NO. 46;
Matches 7; Conservative 6; Mismatches 7; Indels 4; Gaps 1;

OY  1 SRVPTA---WYQHLRLRYRRFV 20
    : ||||| : : : :
Db  243 AEVPTAGRDIMWHEVEKYPNYL 266

RESULT 12
MGDI_MOUSE STANDARD: PRT; 775 AA.
ID MGDI_MOUSE
AC Q9QYH6; Q99PB5; Q9CYX1;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Melanoma-associated antigen DI (MAGE-DI antigen) (Neurotrophin
DE receptor-interacting MAGE homolog) (Dlx1n-1).
GN MAGED1 OR NRAGE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=21265065; PubMed=11084035;
RA Masuda Y., Sasaki A., Shibuya H., Ueno N., Ikeda K., Watanabe K.;
```

```

RT  "Dlx1n-1, a novel protein that binds Dlx5 and regulates its
RT  transcriptional function.";
RL  J. Biol. Chem. 276:5331-5338(2001).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Augier P.H., Chomez P.M., De Backer O.R., Bertrand M.J.M.;
RT  "Ten new murine members of the MAGE gene family.";
RT  Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX  MEDLINE=21085660; PubMed=11217851;
RA  Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA  Arakawa T., Hara A., Fukunishi Y., Kono H., Aachi S., Fukuda S.,
RA  Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA  Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA  Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA  Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA  Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA  Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA  Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA  Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA  Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA  Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
RA  Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA  Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA  Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA  Suzuki H., Toyooka K., Wang K.H., Weitz C., Wiltaker C., Wilming L.,
RA  Wyshnar-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA  Hayashizaki Y.;
RT  "Functional annotation of a full-length mouse cDNA collection.";
RL  Nature 409:685-690(2001).
CC  -I- FUNCTION: Involved in the apoptotic response after nerve growth
CC  factor (NGF) binding in neuronal cells. Binds p75NTR and
CC  antagonizes its association with TrkA, inhibits cell cycle
CC  progression, and facilitates p75NTR-mediated apoptosis. May act as
CC  a regulator of the function of Dlx family members (by similarity).
CC  -I- SUBUNIT: INTERACTS WITH DLX5, DLX7 AND MSX2 AND FORMS
CC  HOMODIMERS.
CC  -I- SUBCELLULAR LOCATION: Cytoplasmic. Expression shifts from the
CC  cytoplasm to the plasma membrane upon stimulation with NGF (By
CC  similarity).
CC  -I- TISSUE SPECIFICITY: Ubiquitously expressed in many adult tissues,
CC  except for the spleen. Expressed in osteoblastic and
CC  chondrogenic cell lines and also during embryonic development.
CC  -I- SIMILARITY: CONTAINS 1 MAGE DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AB029448; BAA87959.1; -
DR  EMBL; AF319975; AAK01203.1; -
DR  EMBL; AK017275; BAB30666.1; -
DR  EMBL; AK013231; BAB28729.1; ALT_INIT.
DR  MGI; MGI:1930187; Maged1.
DR  InterPro; IPR002190; MAGE.
DR  Pfam; PF01454; MAGE; 1.
DR  PROSITE; PS50838; MAGE; 1.
KW  Antigen; Multigene family; Repeat.
FT  DOMAIN 292 441
FT  REPEAT 292 297
FT  REPEAT 298 303
FT  REPEAT 304 309
FT  REPEAT 329 334
FT  REPEAT 335 340
FT  REPEAT 341 346
FT  REPEAT 347 352
```



```

FT REPEAT 353 358 8.
FT REPEAT 359 364 9.
FT REPEAT 365 370 10.
FT REPEAT 371 376 11.
FT REPEAT 377 382 12.
FT REPEAT 383 388 13.
FT REPEAT 389 394 14.
FT REPEAT 395 400 15.
FT REPEAT 401 406 16.
FT REPEAT 407 412 17.
FT REPEAT 413 418 18.
FT REPEAT 419 424 19.
FT REPEAT 425 429 20.
FT REPEAT 430 435 21.
FT REPEAT 436 441 22.
FT DOMAIN 468 666
SQ SEQUENCE 775 AA; 85669 MW; 224B82470816835A CRC64;

```

```

Query Match 31.0%; Score 44; DB 1; Length 775;
Best Local Similarity 42.1%; Pred. No. 53;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

Oy 1 SRVPTAWYQHDLRAYRRF 19
Db 725 SRPFTWARYHONARSRF 743

```

```

RESULT 13
MGDI RAT STANDARD; PRT; 775 AA.
AC 09ES73: 09QX92: 09JH26;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Melanoma-associated antigen DI (MAGE-DI antigen) (Neurotrophin
DE receptor-interacting MAGE homolog) (Sertoli cell neudin
DE related gene-1) (SNERG-1).
GN MAGDI OR NRAGE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid-10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Neural crest;
RX MEDLINE=20439481; PubMed=10985348;
RA Salehi A.H., Roux P.P., Kuba C.J., Zeindler C., Bhakar A.,
RA Tanni L.-L., Verdi J.M., Barker P.A.;
RT *NRAGE, a novel MAGE protein, interacts with the p75 neurotrophin
RT receptor and facilitates nerve growth factor dependent apoptosis.";
RL Neuron 27:279-286(2000).
RN [2]
RP SEQUENCE OF 206-775 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=20466167; PubMed=11014239;
RA Hennry B., Reiter E., Cornet A., Bruyninx M., Daukandt M., Houssa P.,
RA N'Guyen V.-H., Closter J., Hennen G.;
RT *A novel messenger ribonucleic acid homologous to human MAGE-D is
RT strongly expressed in rat Sertoli cells and weakly in Leydig cells
RT and is regulated by folliculin, luteal phase, and prolactin.";
RL Endocrinology 141:3821-3831(2000).
RN [3]
RP SEQUENCE OF 178-775 FROM N.A.
RA Zhang C., He F.;
RT *Rattus norvegicus mRNA for SNERG-1 protein, partial cds.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: Involved in the apoptotic response after nerve growth
CC factor (NGF) binding in neuronal cells. Binds p75NTR and
CC antagonizes its association with TrkA, inhibits cell cycle
CC progression, and facilitates p75NTR-mediated apoptosis. May act as
CC a regulator of the function of DLX family members.

```

```

CC -1- SUBUNIT: INTERACTS WITH DLX5, DLX7 AND MSX2 AND FORMS
CC HOMOMULTIMERS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Expression shifts from the
CC cytoplasm to the plasma membrane upon stimulation with NGF.
CC -1- TISSUE SPECIFICITY: Ubiquitous and in the semiliferous tubules
CC expressed in Sertoli cells but not in germ cells. Expression
CC decreases in all tissues with increased age and is detectable only
CC in brain cortex and lung.
CC -1- DEVELOPMENTAL STAGE: Expressed at low levels throughout the embryo
CC and is enriched in the developing brain and spinal cord.
CC -1- INDUCTION: Folliculin decreased expression while luteal phase and
CC prolactin stimulated expression.
CC -1- SIMILARITY: CONTAINS 1 MAGE DOMAIN.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 726.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF217964; MAG09705.1; -
DR EMBL: AJ133038; CAB65381.1; ALT_FRAME.
DR EMBL: AF274043; AAF75283.1; -
DR InterPro: IPR002190; MAGE.
DR Pfam: PF01454; MAGE; 1.
DR PROSITE: PS50838; MAGE; 1.
DR Antigen: Multigene family; Repeat.
KW DOMAIN 293 441
FT FT 293 441
FT REPEAT 293 298
FT REPEAT 299 304
FT REPEAT 305 310
FT REPEAT 329 334
FT REPEAT 335 340
FT REPEAT 341 346
FT REPEAT 347 352
FT REPEAT 353 358
FT REPEAT 359 364
FT REPEAT 365 370
FT REPEAT 371 376
FT REPEAT 377 382
FT REPEAT 383 388
FT REPEAT 389 394
FT REPEAT 395 400
FT REPEAT 401 406
FT REPEAT 407 412
FT REPEAT 413 418
FT REPEAT 419 424
FT REPEAT 425 429
FT REPEAT 430 435
FT REPEAT 436 441
FT REPEAT 442 447
FT REPEAT 448 453
FT REPEAT 454 459
FT REPEAT 460 465
FT REPEAT 466 471
FT REPEAT 472 477
FT REPEAT 478 483
FT REPEAT 484 489
FT REPEAT 490 495
FT REPEAT 496 501
FT REPEAT 502 507
FT REPEAT 508 513
FT REPEAT 514 519
FT REPEAT 520 525
FT REPEAT 526 531
FT REPEAT 532 537
FT REPEAT 538 543
FT REPEAT 544 549
FT REPEAT 550 555
FT REPEAT 556 561
FT REPEAT 562 567
FT REPEAT 568 573
FT REPEAT 574 579
FT REPEAT 580 585
FT REPEAT 586 591
FT REPEAT 592 597
FT REPEAT 598 603
FT REPEAT 604 609
FT REPEAT 610 615
FT REPEAT 616 621
FT REPEAT 622 627
FT REPEAT 628 633
FT REPEAT 634 639
FT REPEAT 640 645
FT REPEAT 646 651
FT REPEAT 652 657
FT REPEAT 658 663
FT REPEAT 664 669
FT REPEAT 670 675
FT REPEAT 676 681
FT REPEAT 682 687
FT REPEAT 688 693
FT REPEAT 694 699
FT REPEAT 700 705
FT REPEAT 706 711
FT REPEAT 712 717
FT REPEAT 718 723
FT REPEAT 724 729
FT REPEAT 730 735
FT REPEAT 736 741
FT REPEAT 742 747
FT REPEAT 748 753
FT REPEAT 754 759
FT REPEAT 760 765
FT REPEAT 766 771
FT REPEAT 772 777
FT REPEAT 778 783
FT REPEAT 784 789
FT REPEAT 790 795
FT REPEAT 796 801
FT REPEAT 802 807
FT REPEAT 808 813
FT REPEAT 814 819
FT REPEAT 820 825
FT REPEAT 826 831
FT REPEAT 832 837
FT REPEAT 838 843
FT REPEAT 844 849
FT REPEAT 850 855
FT REPEAT 856 861
FT REPEAT 862 867
FT REPEAT 868 873
FT REPEAT 874 879
FT REPEAT 880 885
FT REPEAT 886 891
FT REPEAT 892 897
FT REPEAT 898 903
FT REPEAT 904 909
FT REPEAT 910 915
FT REPEAT 916 921
FT REPEAT 922 927
FT REPEAT 928 933
FT REPEAT 934 939
FT REPEAT 940 945
FT REPEAT 946 951
FT REPEAT 952 957
FT REPEAT 958 963
FT REPEAT 964 969
FT REPEAT 970 975
FT REPEAT 976 981
FT REPEAT 982 987
FT REPEAT 988 993
FT REPEAT 994 999
FT REPEAT 1000 1005
FT REPEAT 1006 1011
FT REPEAT 1012 1017
FT REPEAT 1018 1023
FT REPEAT 1024 1029
FT REPEAT 1030 1035
FT REPEAT 1036 1041
FT REPEAT 1042 1047
FT REPEAT 1048 1053
FT REPEAT 1054 1059
FT REPEAT 1060 1065
FT REPEAT 1066 1071
FT REPEAT 1072 1077
FT REPEAT 1078 1083
FT REPEAT 1084 1089
FT REPEAT 1090 1095
FT REPEAT 1096 1101
FT REPEAT 1102 1107
FT REPEAT 1108 1113
FT REPEAT 1114 1119
FT REPEAT 1120 1125
FT REPEAT 1126 1131
FT REPEAT 1132 1137
FT REPEAT 1138 1143
FT REPEAT 1144 1149
FT REPEAT 1150 1155
FT REPEAT 1156 1161
FT REPEAT 1162 1167
FT REPEAT 1168 1173
FT REPEAT 1174 1179
FT REPEAT 1180 1185
FT REPEAT 1186 1191
FT REPEAT 1192 1197
FT REPEAT 1198 1203
FT REPEAT 1204 1209
FT REPEAT 1210 1215
FT REPEAT 1216 1221
FT REPEAT 1222 1227
FT REPEAT 1228 1233
FT REPEAT 1234 1239
FT REPEAT 1240 1245
FT REPEAT 1246 1251
FT REPEAT 1252 1257
FT REPEAT 1258 1263
FT REPEAT 1264 1269
FT REPEAT 1270 1275
FT REPEAT 1276 1281
FT REPEAT 1282 1287
FT REPEAT 1288 1293
FT REPEAT 1294 1299
FT REPEAT 1300 1305
FT REPEAT 1306 1311
FT REPEAT 1312 1317
FT REPEAT 1318 1323
FT REPEAT 1324 1329
FT REPEAT 1330 1335
FT REPEAT 1336 1341
FT REPEAT 1342 1347
FT REPEAT 1348 1353
FT REPEAT 1354 1359
FT REPEAT 1360 1365
FT REPEAT 1366 1371
FT REPEAT 1372 1377
FT REPEAT 1378 1383
FT REPEAT 1384 1389
FT REPEAT 1390 1395
FT REPEAT 1396 1401
FT REPEAT 1402 1407
FT REPEAT 1408 1413
FT REPEAT 1414 1419
FT REPEAT 1420 1425
FT REPEAT 1426 1431
FT REPEAT 1432 1437
FT REPEAT 1438 1443
FT REPEAT 1444 1449
FT REPEAT 1450 1455
FT REPEAT 1456 1461
FT REPEAT 1462 1467
FT REPEAT 1468 1473
FT REPEAT 1474 1479
FT REPEAT 1480 1485
FT REPEAT 1486 1491
FT REPEAT 1492 1497
FT REPEAT 1498 1503
FT REPEAT 1504 1509
FT REPEAT 1510 1515
FT REPEAT 1516 1521
FT REPEAT 1522 1527
FT REPEAT 1528 1533
FT REPEAT 1534 1539
FT REPEAT 1540 1545
FT REPEAT 1546 1551
FT REPEAT 1552 1557
FT REPEAT 1558 1563
FT REPEAT 1564 1569
FT REPEAT 1570 1575
FT REPEAT 1576 1581
FT REPEAT 1582 1587
FT REPEAT 1588 1593
FT REPEAT 1594 1599
FT REPEAT 1600 1605
FT REPEAT 1606 1611
FT REPEAT 1612 1617
FT REPEAT 1618 1623
FT REPEAT 1624 1629
FT REPEAT 1630 1635
FT REPEAT 1636 1641
FT REPEAT 1642 1647
FT REPEAT 1648 1653
FT REPEAT 1654 1659
FT REPEAT 1660 1665
FT REPEAT 1666 1671
FT REPEAT 1672 1677
FT REPEAT 1678 1683
FT REPEAT 1684 1689
FT REPEAT 1690 1695
FT REPEAT 1696 1701
FT REPEAT 1702 1707
FT REPEAT 1708 1713
FT REPEAT 1714 1719
FT REPEAT 1720 1725
FT REPEAT 1726 1731
FT REPEAT 1732 1737
FT REPEAT 1738 1743
FT REPEAT 1744 1749
FT REPEAT 1750 1755
FT REPEAT 1756 1761
FT REPEAT 1762 1767
FT REPEAT 1768 1773
FT REPEAT 1774 1779
FT REPEAT 1780 1785
FT REPEAT 1786 1791
FT REPEAT 1792 1797
FT REPEAT 1798 1803
FT REPEAT 1804 1809
FT REPEAT 1810 1815
FT REPEAT 1816 1821
FT REPEAT 1822 1827
FT REPEAT 1828 1833
FT REPEAT 1834 1839
FT REPEAT 1840 1845
FT REPEAT 1846 1851
FT REPEAT 1852 1857
FT REPEAT 1858 1863
FT REPEAT 1864 1869
FT REPEAT 1870 1875
FT REPEAT 1876 1881
FT REPEAT 1882 1887
FT REPEAT 1888 1893
FT REPEAT 1894 1899
FT REPEAT 1900 1905
FT REPEAT 1906 1911
FT REPEAT 1912 1917
FT REPEAT 1918 1923
FT REPEAT 1924 1929
FT REPEAT 1930 1935
FT REPEAT 1936 1941
FT REPEAT 1942 1947
FT REPEAT 1948 1953
FT REPEAT 1954 1959
FT REPEAT 1960 1965
FT REPEAT 1966 1971
FT REPEAT 1972 1977
FT REPEAT 1978 1983
FT REPEAT 1984 1989
FT REPEAT 1990 1995
FT REPEAT 1996 2001
FT REPEAT 2002 2007
FT REPEAT 2008 2013
FT REPEAT 2014 2019
FT REPEAT 2020 2025
FT REPEAT 2026 2031
FT REPEAT 2032 2037
FT REPEAT 2038 2043
FT REPEAT 2044 2049
FT REPEAT 2050 2055
FT REPEAT 2056 2061
FT REPEAT 2062 2067
FT REPEAT 2068 2073
FT REPEAT 2074 2079
FT REPEAT 2080 2085
FT REPEAT 2086 2091
FT REPEAT 2092 2097
FT REPEAT 2098 2103
FT REPEAT 2104 2109
FT REPEAT 2110 2115
FT REPEAT 2116 2121
FT REPEAT 2122 2127
FT REPEAT 2128 2133
FT REPEAT 2134 2139
FT REPEAT 2140 2145
FT REPEAT 2146 2151
FT REPEAT 2152 2157
FT REPEAT 2158 2163
FT REPEAT 2164 2169
FT REPEAT 2170 2175
FT REPEAT 2176 2181
FT REPEAT 2182 2187
FT REPEAT 2188 2193
FT REPEAT 2194 2199
FT REPEAT 2200 2205
FT REPEAT 2206 2211
FT REPEAT 2212 2217
FT REPEAT 2218 2223
FT REPEAT 2224 2229
FT REPEAT 2230 2235
FT REPEAT 2236 2241
FT REPEAT 2242 2247
FT REPEAT 2248 2253
FT REPEAT 2254 2259
FT REPEAT 2260 2265
FT REPEAT 2266 2271
FT REPEAT 2272 2277
FT REPEAT 2278 2283
FT REPEAT 2284 2289
FT REPEAT 2290 2295
FT REPEAT 2296 2301
FT REPEAT 2302 2307
FT REPEAT 2308 2313
FT REPEAT 2314 2319
FT REPEAT 2320 2325
FT REPEAT 2326 2331
FT REPEAT 2332 2337
FT REPEAT 2338 2343
FT REPEAT 2344 2349
FT REPEAT 2350 2355
FT REPEAT 2356 2361
FT REPEAT 2362 2367
FT REPEAT 2368 2373
FT REPEAT 2374 2379
FT REPEAT 2380 2385
FT REPEAT 2386 2391
FT REPEAT 2392 2397
FT REPEAT 2398 2403
FT REPEAT 2404 2409
FT REPEAT 2410 2415
FT REPEAT 2416 2421
FT REPEAT 2422 2427
FT REPEAT 2428 2433
FT REPEAT 2434 2439
FT REPEAT 2440 2445
FT REPEAT 2446 2451
FT REPEAT 2452 2457
FT REPEAT 2458 2463
FT REPEAT 2464 2469
FT REPEAT 2470 2475
FT REPEAT 2476 2481
FT REPEAT 2482 2487
FT REPEAT 2488 2493
FT REPEAT 2494 2499
FT REPEAT 2500 2505
FT REPEAT 2506 2511
FT REPEAT 2512 2517
FT REPEAT 2518 2523
FT REPEAT 2524 2529
FT REPEAT 2530 2535
FT REPEAT 2536 2541
FT REPEAT 2542 2547
FT REPEAT 2548 2553
FT REPEAT 2554 2559
FT REPEAT 2560 2565
FT REPEAT 2566 2571
FT REPEAT 2572 2577
FT REPEAT 2578 2583
FT REPEAT 2584 2589
FT REPEAT 2590 2595
FT REPEAT 2596 2601
FT REPEAT 2602 2607
FT REPEAT 2608 2613
FT REPEAT 2614 2619
FT REPEAT 2620 2625
FT REPEAT 2626 2631
FT REPEAT 2632 2637
FT REPEAT 2638 2643
FT REPEAT 2644 2649
FT REPEAT 2650 2655
FT REPEAT 2656 2661
FT REPEAT 2662 2667
FT REPEAT 2668 2673
FT REPEAT 2674 2679
FT REPEAT 2680 2685
FT REPEAT 2686 2691
FT REPEAT 2692 2697
FT REPEAT 2698 2703
FT REPEAT 2704 2709
FT REPEAT 2710 2715
FT REPEAT 2716 2721
FT REPEAT 2722 2727
FT REPEAT 2728 2733
FT REPEAT 2734 2739
FT REPEAT 2740 2745
FT REPEAT 2746 2751
FT REPEAT 2752 2757
FT REPEAT 2758 2763
FT REPEAT 2764 2769
FT REPEAT 2770 2775
FT REPEAT 2776 2781
FT REPEAT 2782 2787
FT REPEAT 2788 2793
FT REPEAT 2794 2799
FT REPEAT 2800 2805
FT REPEAT 2806 2811
FT REPEAT 2812 2817
FT REPEAT 2818 2823
FT REPEAT 2824 2829
FT REPEAT 2830 2835
FT REPEAT 2836 2841
FT REPEAT 2842 2847
FT REPEAT 2848 2853
FT REPEAT 2854 2859
FT REPEAT 2860 2865
FT REPEAT 2866 2871
FT REPEAT 2872 2877
FT REPEAT 2878 2883
FT REPEAT 2884 2889
FT REPEAT 2890 2895
FT REPEAT 2896 2901
FT REPEAT 2902 2907
FT REPEAT 2908 2913
FT REPEAT 2914 2919
FT REPEAT 2920 2925
FT REPEAT 2926 2931
FT REPEAT 2932 2937
FT REPEAT 2938 2943
FT REPEAT 2944 2949
FT REPEAT 2950 2955
FT REPEAT 2956 2961
FT REPEAT 2962 2967
FT REPEAT 2968 2973
FT REPEAT 2974 2979
FT REPEAT 2980 2985
FT REPEAT 2986 2991
FT REPEAT 2992 2997
FT REPEAT 2998 3003
FT REPEAT 3004 3009
FT REPEAT 3010 3015
FT REPEAT 3016 3021
FT REPEAT 3022 3027
FT REPEAT 3028 3033
FT REPEAT 3034 3039
FT REPEAT 3040 3045
FT REPEAT 3046 3051
FT REPEAT 3052 3057
FT REPEAT 3058 3063
FT REPEAT 3064 3069
FT REPEAT 3070 3075
FT REPEAT 3076 3081
FT REPEAT 3082 3087
FT REPEAT 3088 3093
FT REPEAT 3094 3099
FT REPEAT 3100 3105
FT REPEAT 3106 3111
FT REPEAT 3112 3117
FT REPEAT 3118 3123
FT REPEAT 3124 3129
FT REPEAT 3130 3135
FT REPEAT 3136 3141
FT REPEAT 3142 3147
FT REPEAT 3148 3153
FT REPEAT 3154 3159
FT REPEAT 3160 3165
FT REPEAT 3166 3171
FT REPEAT 3172 3177
FT REPEAT 3178 3183
FT REPEAT 3184 3189
FT REPEAT 3190 3195
FT REPEAT 3196 3201
FT REPEAT 3202 3207
FT REPEAT 3208 3213
FT REPEAT 3214 3219
FT REPEAT 3220 3225
FT REPEAT 3226 3231
FT REPEAT 3232 3237
FT REPEAT 3238 3243
FT REPEAT 3244 3249
FT REPEAT 3250 3255
FT REPEAT 3256 3261
FT REPEAT 3262 3267
FT REPEAT 3268 3273
FT REPEAT 3274 3279
FT REPEAT 3280 3285
FT REPEAT 3286 3291
FT REPEAT 3292 3297
FT REPEAT 3298 3303
FT REPEAT 3304 3309
FT REPEAT 3310 3315
FT REPEAT 3316 3321
FT REPEAT 3322 3327
FT REPEAT 3328 3333
FT REPEAT 3334 3339
FT REPEAT 3340 3345
FT REPEAT 3346 3351
FT REPEAT 3352 3357
FT REPEAT 3358 3363
FT REPEAT 3364 3369
FT REPEAT 3370 3375
FT REPEAT 3376 3381
FT REPEAT 3382 3387
FT REPEAT 3388 3393
FT REPEAT 3394 3399
FT REPEAT 3400 3405
FT REPEAT 3406 3411
FT REPEAT 3412 3417
FT REPEAT 3418 3423
FT REPEAT 3424 3429
FT REPEAT 3430 3435
FT REPEAT 3436 3441
FT REPEAT 3442 3447
FT REPEAT 3448 3453
FT REPEAT 3454 3459
FT REPEAT 3460 3465
FT REPEAT 3466 3471
FT REPEAT 3472 3477
FT REPEAT 3478 3483
FT REPEAT 3484 3489
FT REPEAT 3490 3495
FT REPEAT 3496 3501
FT REPEAT 3502 3507
FT REPEAT 3508 3513
FT REPEAT 3514 3519
FT REPEAT 3520 3525
FT REPEAT 3526 3531
FT REPEAT 3532 3537
FT REPEAT 3538 3543
FT REPEAT 3544 3549
FT REPEAT 3550 3555
FT REPEAT 3556 3561
FT REPEAT 3562 3567
FT REPEAT 3568 3573
FT REPEAT 3574 3579
FT REPEAT 3580 3585
FT REPEAT 3586 3591
FT REPEAT 3592 3597
FT REPEAT 3598 3603
FT REPEAT 3604 3609
FT REPEAT 3610 3615
FT REPEAT 3616 3621
FT REPEAT 3622 3627
FT REPEAT 3628 3633
FT REPEAT 3634 3639
FT REPEAT 3640 3645
FT REPEAT 3646 3651
FT REPEAT 3652 3657
FT REPEAT 3658 3663
FT REPEAT 3664 3669
FT REPEAT 3670 3675
FT REPEAT 3676 3681
FT REPEAT 3682 3687
FT REPEAT 3688 3693
FT REPEAT 3694 3699
FT REPEAT 3700 3705
FT REPEAT 3706 3711
FT REPEAT 3712 3717
FT REPEAT 3718 3723
FT REPEAT 3724 3729
FT REPEAT 3730 3735
FT REPEAT 3736 3741
FT REPEAT 3742 3747
FT REPEAT 3748 3753
FT REPEAT 3754 3759
FT REPEAT 3760 3765
FT REPEAT 3766 3771
FT REPEAT 3772 3777
FT REPEAT 3778 3783
FT REPEAT 3784 3789
FT REPEAT 3790 3795
FT REPEAT 3796 3801
FT REPEAT 3802 3807
FT REPEAT 3808 3813
FT REPEAT 3814 3819
FT REPEAT 3820 3825
FT REPEAT 3826 3831
FT REPEAT 3832 3837
FT REPEAT 3838 3843
FT REPEAT 3844 3849
FT REPEAT 3850 3855
FT REPEAT 3856 3861
FT REPEAT 3862 3867
FT REPEAT 3868 3873
FT REPEAT 3874 3879
FT REPEAT 3880 3885
FT REPEAT 3886 3891
FT REPEAT 3892 3897
FT REPEAT 3898 3903
FT REPEAT 3904 3909
FT REPEAT 3910 3915
FT REPEAT 3916 3921
FT REPEAT 3922 3927
FT REPEAT 3928 3933
FT REPEAT 3934 3939
FT REPEAT 3940 3945
FT REPEAT 3946 3951
FT REPEAT 3952 3957
FT REPEAT 3958 3963
FT REPEAT 3964 3969
FT REPEAT 3970 3975
FT REPEAT 3976 3981
FT REPEAT 3982 3987
FT REPEAT 3988 3993
FT REPEAT 3994 3999
FT REPEAT 4000 4005
FT REPEAT 4006 4011
FT REPEAT 4012 4017
FT REPEAT 4018 4023
FT REPEAT 4024 4029
FT REPEAT 4030 4035
FT REPEAT 4036 4041
FT REPEAT 4042 4047
FT REPEAT 4048 4053
FT REPEAT 4054 4059
FT REPEAT 4060 4065
FT REPEAT 4066 4071
FT REPEAT 4072 4077
FT REPEAT 4078 4083
FT REPEAT 4084 4089
FT REPEAT 4090 4095
FT REPEAT 4096 4101
FT REPEAT 4102 4107
FT REPEAT 4108 4113
FT REPEAT 4114 4119
FT REPEAT 4120 4125
FT REPEAT 4126 4131
FT REPEAT 4132 4137
FT REPEAT 4138 4143
FT REPEAT 4144 4149
FT REPEAT 4150 4155
FT REPEAT 4156 4161
FT REPEAT 4162 4167
FT REPEAT 4168 4173
FT REPEAT 4174 4179
FT REPEAT 4180 4185
FT REPEAT 4186 4191
FT REPEAT 4192 4197
FT REPEAT 4198 4203
FT REPEAT 4204 4209
FT REPEAT 4210 4215
FT REPEAT 4216 4221
FT REPEAT 4222 4227
FT REPEAT 4228 4233
FT REPEAT 4234 4239
FT REPEAT 4240 4245
FT REPEAT 4246 4251
FT REPEAT 4252 4257
FT REPEAT 4258 4263
FT REPEAT 4264 4269
FT REPEAT 4270 4275
FT REPEAT 4276 4281
FT REPEAT 4282 4287
FT REPEAT 4288 4293
FT REPEAT 4294 4299
FT REPEAT 4300 4305
FT REPEAT 4306 4311
FT REPEAT 4312 4317
FT REPEAT 4318 4323
FT REPEAT 4324 4329
FT REPEAT 4330 4335
FT REPEAT 4336 4341
FT REPEAT 4342 4347
FT REPEAT 4348 4353
FT REPEAT 4354 4359
FT REPEAT 4360 4365
FT REPEAT 4366 4371
FT REPEAT 4372 4377
FT REPEAT 4378 4383
FT REPEAT 4384 4389
FT REPEAT 4390 4395
FT REPEAT 4396 4401
FT REPEAT 4402 4407
FT REPEAT 4408 4413
FT REPEAT 4414 4419
FT REPEAT 4420 4425
FT REPEAT 4426 4431
FT REPEAT 4432 4437
FT REPEAT 4438 4443
FT REPEAT 4444 4449
FT REPEAT 4450 4455
FT REPEAT 4456 4461
FT REPEAT 4462 4467
FT REPEAT 4468 4473
FT REPEAT 4474 4479
FT REPEAT 4480 4485
FT REPEAT 4486 4491
FT REPEAT 4492 4497
FT REPEAT 4498 4503
FT REPEAT 4504 4509
FT REPEAT 4510 4515
FT REPEAT 4516 4521
FT REPEAT 4522 4527
FT REPEAT 4528 4533
FT REPEAT 4534 4539
FT REPEAT 4540 4545
FT REPEAT 4546 4551
FT REPEAT 4552 4557
FT REPEAT 4558 4563
FT REPEAT 4564 4569
FT REPEAT 4570 4575
FT REPEAT 4576 4581
FT REPEAT 4582 4587
FT REPEAT 4588 4593
FT REPEAT 4594 4599
FT REPEAT 4600 4605
FT REPEAT 4606 4611
FT REPEAT 4612 4617
FT REPEAT 4618 4623
FT REPEAT 4624 4629
FT REPEAT 4630 4635
FT REPEAT 4636 4641
FT REPEAT 4642 4647
FT REPEAT 4648 4653
FT REPEAT 4654 4659
FT REPEAT 4660 4665
FT REPEAT 4666 4671
FT REPEAT 4672 4677
FT REPEAT 4678 4683
FT REPEAT 4684 4689
FT REPEAT 4690 4695
FT REPEAT 4696 4701
FT REPEAT 4702 4707
FT REPEAT 4708 4713
FT REPEAT 4714 4719
FT REPEAT 4720 4725
FT REPEAT 4726 4731
FT REPEAT 4732 4737
FT REPEAT 4738 4743
FT REPEAT 4744 4749
FT REPEAT 4750 4755
FT REPEAT 4756 4761
FT REPEAT 4762 4767
FT REPEAT 4768 4773
FT REPEAT 4774 4779
FT REPEAT 4780 4785
FT REPEAT 4786 4791
FT REPEAT 4792 4797
FT REPEAT 4798 4803
FT REPEAT 4804 4809
FT REPEAT 4810 4815
FT REPEAT 4816 4821
FT REPEAT 4822 4827
FT REPEAT 4828 4833
FT REPEAT 4834 4839
FT REPEAT 4840 4845
FT REPEAT 4846 4851
FT REPEAT 4852
```

MGDI\_HUMAN STANDARD; PRT; 778 AA.  
 ID MGDI\_HUMAN  
 AC Q9YV3; Q9UE36; Q9HBT4; Q9H352;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Melanoma-associated antigen DI (MAGE-DI antigen) (Neurotrophin receptor-interacting MAGE homolog) (PRO2292).  
 GN MAGE1 OR NRAGE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=99339980; PubMed=10409427;  
 RA Pold M., Zhou J., Chen G.L., Hall J.M., Vescio R.A., Berenson J.R.;  
 RT "Identification of a new, unorthodox member of the MAGE gene family."; Genomics 59:161-167(1999).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RX Salehi A.H., Roux P.P., Kudu C.J., Zeindler C., Bhakar A.,  
 RA Tannis L.L., Verdi J.M., Barker P.A.;  
 RT "NRAGE, a novel MAGE protein, interacts with the p75 neurotrophin receptor and facilitates nerve growth factor dependent apoptosis."; Neuron 27:279-288(2000).  
 RL [3]  
 RP SEQUENCE OF 304-778 FROM N.A.  
 RC TISSUE-Testis;  
 RA Blum H., Bauersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 396-778 FROM N.A.  
 RC TISSUE-Fetal liver;  
 RX MEDLINE=21177478; PubMed=11280991;  
 RA Zhang C.G., Xing G.C., Wei H.D., Yu Y.T., He F.C.;  
 RT "A new melanoma antigen-encoding gene subfamily in human chromosome X";  
 RL I Chuan Hsueh Pao 28:197-203(2001).  
 RN [5]  
 RP IDENTIFICATION OF THE TRANSLATIONAL INITIATION CODON.  
 RX MEDLINE=20541720; PubMed=11087672;  
 RA Kudu C.J., Goldhawk D.G., Barker P.A., Verdi J.M.;  
 RT "Identification of the translational initiation codon in human MAGE1."; Genomics 70:150-152(2000).  
 RL [6]  
 RP FUNCTION: Involved in the apoptotic response after nerve growth factor (NGF) binding in neuronal cells. Inhibits p75NTR and antagonizes its association with TRKA, inhibits cell cycle progression, and facilitates p75NTR-mediated apoptosis. May act as a regulator of the function of DLX family members.  
 CC -1 SUBUNIT: Interacts with the p75 neurotrophin receptor.  
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic. Expression shifts from the cytoplasm to the plasma membrane upon stimulation with NGF (By similarity).  
 CC -1 TISSUE SPECIFICITY: EXPRESSED IN BONE MARROW STROMAL CELLS FROM BOTH MULTIPLE MYELOMA PATIENTS AND HEALTHY DONORS. SEEMS TO BE UNIGENOUSLY EXPRESSED.  
 CC -1 SIMILARITY: CONTAINS 1 MAGE DOMAIN.  
 CC -1 CAUTION: REF.1 differs from that shown due to several frameshifts that resulted in a N-terminally truncated protein.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AF124440; AAD31421.1; ALT\_FRAME.

DR EMBL; AF217963; AAG09704.1; -;  
 DR EMBL; AL133628; CAB63752.1; -;  
 DR EMBL; AF132205; AAG35551.1; ALT\_INIT.  
 DR MIM: 300224; -;  
 DR InterPro: IPR002190; MAGE.  
 DR Pfam: PF01454; MAGE; 1.  
 DR PROSITE: PS0838; MAGE; 1.  
 KW Antigen; Multigene family;  
 FT DOMAIN 296 444  
 FT  
 FT REPEAT 296 301 1. X.  
 FT REPEAT 302 307 2.  
 FT REPEAT 308 313 3.  
 FT REPEAT 332 337 4.  
 FT REPEAT 338 343 5.  
 FT REPEAT 344 349 6.  
 FT REPEAT 350 355 7.  
 FT REPEAT 356 361 8.  
 FT REPEAT 362 367 9.  
 FT REPEAT 368 373 10.  
 FT REPEAT 374 379 11.  
 FT REPEAT 380 385 12.  
 FT REPEAT 386 391 13.  
 FT REPEAT 392 397 14.  
 FT REPEAT 398 403 15.  
 FT REPEAT 404 409 16.  
 FT REPEAT 410 415 17.  
 FT REPEAT 416 421 18.  
 FT REPEAT 422 427 19.  
 FT REPEAT 428 432 20 (IMPERFECT).  
 FT REPEAT 433 438 21.  
 FT REPEAT 439 444 22.  
 FT DOMAIN 471 669  
 SQ SEQUENCE 778 AA; 86150 MW; 0F8BEC7155326FCC CRC64;  
 OY 1 SRVPTAWYOHDLRAYRRF 19  
 DB 728 SRPFTWARYHONARSRF 746  
 Query Match 31.0%; Score 44; DB 1; Length 778;  
 Best Local Similarity 42.1%; Pred. No. 53;  
 Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 RESULT 15  
 PRTND\_HUMAN STANDARD; PRT; 2485 AA.  
 ID PRTND\_HUMAN  
 AC Q12923; Q15263; Q16826; Q15264; Q15265; Q15159;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 13 (BC 3.1.3.48)  
 DE (Protein-tyrosine phosphatase 1E) (PTP-1E) (Protein-tyrosine phosphatase PTP1L) (Fas-associated protein-tyrosine phosphatase 1)  
 DE (FAP-1).  
 GN PTPN13 OR PTP1E OR PTP1L OR PNP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=94350988; PubMed=8071359;  
 RA Banville D., Ahmad S., Stocco R., Shen S.-H.;  
 RT "A novel protein-tyrosine phosphatase with homology to both the cytoskeletal proteins of the band 4.1 family and junction-associated guanylate kinases."; J. Biol. Chem. 269:22320-22327(1994).  
 RL [2]  
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RP TISSUE=Leukemia;

```

RX MEDLINE-94116679; PubMed-8287977;
RA Maekawa K., Imagawa N., Nagamatsu M., Harada S.;
RT "Molecular cloning of a novel protein-tyrosine phosphatase containing
RT a membrane-binding domain and GIGF repeats.";
RL FEBS Lett. 337:200-206(1994).
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RA MEDLINE-95014139; PubMed-7929060;
RX Sara J., Claesson-Welsh L., Heldin C.-H., Gonen L.J.;
RT "Cloning and characterization of PRPL, a protein tyrosine phosphatase
RT with similarities to cytoskeletal-associated proteins.";
RL J. Biol. Chem. 269:24082-24089(1994).
RN
RP SEQUENCE OF 1216-2490 FROM N.A.
RC TISSUE=Pancreas;
RA Wang H.Y.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
RN
RP STRUCTURE BY NMR OF 1361-1456.
RX MEDLINE-20170882; PubMed-10704206;
RA Kozlov G., Gehring K., Ekiel I.;
RT "Solution structure of the PDZ2 domain from human phosphatase hPTPL
RT and its interactions with C-terminal peptides from the Fas
RT receptor.";
RL Biochemistry 39:2572-2580(2000).
CC -1- FUNCTION: BINDS TO A NEGATIVE REGULATORY DOMAIN IN FAS THAT
CC INHIBITS FAS-INDUCED APOPTOSIS.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES WITH THE EXCEPTION OF
CC THE LIVER AND SKELETAL MUSCLE. MOST ABUNDANT IN LUNG, KIDNEY AND
CC FETAL BRAIN.
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE PDZ/DHR DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12128; AAB60339.1; -
DR EMBL; D12209; BAA04750.1; -
DR EMBL; D12120; BAA04751.1; -
DR EMBL; D12121; BAA04752.1; -
DR EMBL; X80289; CAA56563.1; -
DR EMBL; X79676; CAA56124.1; -
DR PDB; 3PDZ; 17-MAR-00.
DR MIM; 600267; -
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_prot_phphatase.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00595; PDZ; 5.
DR Pfam; PR00102; Y_phosphatase; 1.
DR PRINTS; PR00935; BAND41.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00295; B41; 1.
DR SMART; SM00228; PDZ; 5.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00660; BAND_41_1; FALSE_NEG.
DR PROSITE; PS00661; BAND_41_2; FALSE_NEG.
DR PROSITE; PS0057; BAND_41_3; 1.
DR PROSITE; PS0106; PDZ; 5.

```

```

DR PROSITE; PS00383; TYR_PHOSPHATASE_1; FALSE_NEG.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Structural protein; Cytoskeleton; Hydrolase; Repeat; 3D-structure;
KW Alternative splicing; Coiled coil.
FT DOMAIN 56 59 POLY-LEU.
FT DOMAIN 585 879 BAND 4.1-LIKE.
FT DOMAIN 2237 2485 PROTEIN-TYROSINE PHOSPHATASE.
FT DOMAIN 379 399 COILED COIL (POTENTIAL).
FT DOMAIN 469 504 COILED COIL (POTENTIAL).
FT DOMAIN 1775 1804 COILED COIL (POTENTIAL).
FT DOMAIN 2057 2085 COILED COIL (POTENTIAL).
FT DOMAIN 1093 1178 PDZ 1.
FT DOMAIN 1368 1452 PDZ 2.
FT DOMAIN 1501 1588 PDZ 3.
FT DOMAIN 1788 1868 PDZ 4.
FT DOMAIN 1882 1965 PDZ 5.
FT DOMAIN 1742 1749 POLY-SER.
FT ACT_SITE 2408 2408 BY SIMILARITY.
FT VARSPPLIC 884 1074 MISSING (IN ISOFORM 2).
FT VARSPPLIC 1056 1074 MISSING (IN ISOFORM 3).
FT CONFLICT 1134 1135 LD -> FH (IN REF. 3).
FT CONFLICT 1216 1229 KDHMSRGTLRHIS -> DLSRSHCVLAHL (IN
FT REF. 4).
FT CONFLICT 1238 1239 GL -> A (IN REF. 4).
FT CONFLICT 1357 1357 S -> P (IN REF. 4).
FT CONFLICT 1362 1363 K -> RS (IN REF. 4).
FT CONFLICT 1383 1383 T -> TVLFDK (IN REF. 1).
FT CONFLICT 1538 1538 P -> A (IN REF. 3).
FT CONFLICT 1649 1649 R -> K (IN REF. 4).
FT CONFLICT 1698 1714 KSEDICICMEFYPPQKI -> RVKKIPFVPCFTILKKR
FT (IN REF. 4).
FT CONFLICT 1797 1797 G -> A (IN REF. 3).
FT CONFLICT 1856 1857 AA -> G (IN REF. 4).
FT CONFLICT 2069 2069 A -> S (IN REF. 4).
FT CONFLICT 2206 2210 GLDQ -> VARS (IN REF. 4).
SQ SEQUENCE 2485 AA; 276903 MW; 8D1B31597C66962B CRC64;

```

```

Query Match 31.0%; Score 44; DB 1; Length 2485;
Best Local Similarity 31.2%; Pred.No.1.8e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 10; Gaps 1;
OY 1 SRVPTAWYQHD-----LRAVRFVNR 22
DB 2369 SHLFTAMPDHPSPQDDLLTFISTYKRIHR 2400

```

Search completed: August 15, 2002, 12:05:56  
Job time: 371 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 12:05:33 ; Search time 41.42 Seconds  
(Without alignments)  
100.238 Million cell updates/sec

Title: DEVI-613-COMBFRRG  
Perfect score: 142  
Sequence: 1 SRVPTAWYQHDLRAYRRFVHRP 24

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	38.0	345	2	09X7U8 streptomyc
2	52.5	37.0	267	2	048706 lactobacill
3	51	35.9	389	16	053357 mycobacteri
4	51	35.9	454	5	09NAK6 caenorhabdi
5	51	35.9	512	5	09V820 drosophila
6	51	35.9	833	16	09JORS neisseria m
7	49.5	34.9	141	16	09ZK25 rhizobium m
8	49	34.5	417	5	076609 caenorhabdi
9	49	34.5	431	10	094CQ0 oryza sativ
10	48.5	34.2	104	2	09AFK4 shigella fl
11	48.5	34.2	434	17	030067 archaeoglob
12	48.5	34.2	506	16	09KOS2 neisseria m
13	48.5	34.2	833	2	09JPE1 neisseria m
14	48	33.8	269	16	09X4D5 streptococ
15	48	33.8	295	16	053514 mycobacteri
16	48	33.8	702	2	09AD11 streptomyc

17	47.5	33.5	247	11	09D123 mus musculu
18	47.5	33.5	417	13	013013 ambystoma m
19	47	33.1	200	2	09ZBU7 streptomyc
20	47	33.1	361	16	09ZLW6 rhizobium m
21	47	33.1	5388	5	09UID0 leishmania
22	46.5	32.7	1186	12	055767 chilo iride
23	46	32.4	119	2	09F3D7 streptococ
24	46	32.4	332	12	082493 influenza a
25	46	32.4	463	10	09FWO3 oryza sativ
26	46	32.0	467	2	09RLI0 streptomyc
27	45.5	32.0	282	5	09W5H9 drosophila
28	45.5	32.0	286	2	051376 pseudomonas
29	45.5	32.0	378	5	09W5H8 drosophila
30	45.5	32.0	382	16	09I0Q0 pseudomonas
31	45.5	32.0	2292	12	066765 encephalomy
32	45.5	32.0	2292	12	066850 encephalomy
33	45	31.7	173	2	004309 salmonella
34	45	31.7	253	16	09RR75 delnoccocus
35	45	31.7	416	16	09I235 pseudomonas
36	45	31.7	568	4	09UN89 homo sapien
37	45	31.7	568	4	095470 homo sapien
38	45	31.7	580	4	09UG68 homo sapien
39	45	31.7	642	2	09LIJ6 streptomyc
40	45	31.7	663	16	09PEJ9 xylella fas
41	45	31.7	754	12	09JH33 tt virus. o
42	45	31.7	790	5	020599 caenorhabdi
43	45	31.7	842	5	0950F5 caenorhabdi
44	45	31.7	1776	10	09MA20 arabidopsis
45	45	31.7	2484	6	028006 bos taurus

#### ALIGNMENTS

RESULT	ID	Query Match	Length	DB ID	Description
1	09X7U8	38.0	345	2	09X7U8 streptomyc
2	09X7U8	37.0	267	2	048706 lactobacill
3	09X7U8	35.9	389	16	053357 mycobacteri
4	09X7U8	35.9	454	5	09NAK6 caenorhabdi
5	09X7U8	35.9	512	5	09V820 drosophila
6	09X7U8	35.9	833	16	09JORS neisseria m
7	09X7U8	34.9	141	16	09ZK25 rhizobium m
8	09X7U8	34.5	417	5	076609 caenorhabdi
9	09X7U8	34.5	431	10	094CQ0 oryza sativ
10	09X7U8	34.2	104	2	09AFK4 shigella fl
11	09X7U8	34.2	434	17	030067 archaeoglob
12	09X7U8	34.2	506	16	09KOS2 neisseria m
13	09X7U8	34.2	833	2	09JPE1 neisseria m
14	09X7U8	33.8	269	16	09X4D5 streptococ
15	09X7U8	33.8	295	16	053514 mycobacteri
16	09X7U8	33.8	702	2	09AD11 streptomyc

Query Match 38.0%; Score 54; DB 2; Length 345;  
Best Local Similarity 45.0%; Pred. No. 5.4;

Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 5 WTAWYQHDLRAYRFPVHRRP 24  
 1 : : : | | | | : | |  
 Db 82 WAERSRHEIGAYRFPVRRP 101

RESULT 2

Q48706 PRELIMINARY; PRT; 267 AA.

AC Q48706;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 GN ORF1 PROTEIN.  
 OS Lactobacillus leichmannii.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Lactobacillus  
 OX NCBI\_TaxID=28039;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 20076;  
 RA Schenk-Groeninger R.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X81869; CAA57458.1; -;  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SO SEQUENCE 267 AA; 29995 MW; 29D06B935C559AB CRC64;

Query Match 37.0%; Score 52.5; DB 2; Length 267;  
 Best Local Similarity 52.4%; Pred. No. 6.9;  
 Matches 11; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

Qy 2 RYPWTAWYQHDLRAY---RRF 19  
 : | | : | | | | | | | | | |  
 Db 161 QVPLAVYTNHDLRYNCCORF 181

RESULT 3

ID O53357 PRELIMINARY; PRT; 389 AA.

AC O53357;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE N-ACYL-L-AMINO ACID AMIDOHYDROLASE.  
 GN AMIA OR RV3305C OR MTY016.04C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RC MEDLINE=96295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL; AL021841; CAA17077.1; -;  
 DR MEROPS; M40. UNM; -;  
 DR Tuberculist; RV3305C; -;  
 DR InterPro; IPR002933; Peptidase\_M20.  
 DR InterPro; IPR00169; ThiolProt\_act\_site.

DR Pfam; PF01546; Peptidase\_M20; 1.  
 DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; UNKNOWN\_1.  
 KN Hydrolase; Complete proteome.  
 SO SEQUENCE 389 AA; 41227 MW; B7D595748A1522A0 CRC64;

Query Match 35.9%; Score 51; DB 16; Length 389;  
 Best Local Similarity 45.0%; Pred. No. 17;  
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 5 WTAWYQHDLRAYRFPVHRRP 24  
 1 : : : | | | | : | |  
 Db 10 WLAAHDDLVGWRRIHRRP 29

RESULT 4

ID Q9NAK6 PRELIMINARY; PRT; 454 AA.

AC Q9NAK6;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Y38E10A.4 PROTEIN.  
 GN Y38E10A.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wallis J.M.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018(1998).  
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
 DR EMBL; AL110484; CAB54395.1; -;  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR001304; Lectin\_c.  
 DR Pfam; PF00431; CUB; 1.  
 DR Pfam; PF00059; Lectin\_c; 2.  
 DR SMART; SM00034; CLECT; 2.  
 DR SMART; SM00042; CUB; 1.  
 DR PROSITE; PS01180; CUB; 1.  
 DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; UNKNOWN\_1.  
 DR PROSITE; PS50041; C\_TYPE\_LECTIN\_2; 2.  
 SO SEQUENCE 454 AA; 50121 MW; 452944C0E1454A39 CRC64;

Query Match 35.9%; Score 51; DB 5; Length 454;  
 Best Local Similarity 41.7%; Pred. No. 19;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 1 SRVPTAWYQHDLRAYRFPVHRRP 24  
 1 : | : | | : | : | : | : |  
 Db 258 SMVPTGSMVYPCVCKASNSFLCKRP 281

RESULT 5

ID Q9V820 PRELIMINARY; PRT; 512 AA.

AC Q9V820; Q9V821;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE BEST:LD13441 PROTEIN.  
 GN BEST:LD13441 OR CG14478.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyroidae; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutten G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Milos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch L.C., Baldwin D.,  
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu K., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier A.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harits N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeter F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Maassman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC EMBL: AE003803; AAF57859.1; -;  
 DR EMBL: AE003803; AAF57860.1; -;  
 DR FlyBase: FBgn0028953; BEST: ID13441.  
 DR InterPro: IPR001525; C5\_DNA\_meth.  
 DR PROSITE: PS00095; C5\_MTase\_2; UNKNOWN\_1.  
 KW Alternative splicing; Hypothetical protein.  
 FT VARSLIC 1 366 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 512 AA; 56005 MW; 943FB812DD3C1B91 CRC64;

Query Match 35.9%; Score 51; DB 5; Length 512;  
 Best Local Similarity 50.0%; Pred. No. 22;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 7 AMYOHDLRAVRYRFR 24  
 Db 340 SYFHDLRYRFR 357

RESULT 6  
 ID 09J0R5 PRELIMINARY; PRT; 833 AA.  
 AC 09J0R5;  
 DT 01-OCT-2000 (TEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TEMBLrel. 15, last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)

DE HYPOTHETICAL PROTEIN NMA0692.  
 GN NMA0692 OR RTM4.  
 OS Neisseria meningitidis (serogroup A), and  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_Taxid=65699, 487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,  
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.,  
 RT "Complete DNA sequence of a serogroup A strain of *Neisseria*  
 RL meningitidis 22491.";  
 RL Nature 404:502-506(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20187481; PubMed=10722605;  
 RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,  
 RA Tansley C.R.;  
 RT "Molecular and biological analysis of eight genetic islands that  
 RT distinguish *neisseria meningitidis* from the closely related pathogen  
 RT *neisseria gonorrhoeae*.";  
 RL Infect. Immun. 68:2082-2095(2000).  
 DR EMBL: AL162753; CAB83978.1; -;  
 DR EMBL: AJ391255; CAB71949.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 833 AA; 88713 MW; DBC91A437511AC9 CRC64;

Query Match 35.9%; Score 51; DB 16; Length 833;  
 Best Local Similarity 33.3%; Pred. No. 36;  
 Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 2 RVPTARYOHDLRAVRYRFR 22  
 Db 666 RTGEPWTRHVDYRQRYNO 686

RESULT 7  
 ID 092K25 PRELIMINARY; PRT; 141 AA.  
 AC 092K25;  
 DT 01-DEC-2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)  
 DE HYPOTHETICAL TRANSMEMBRANE PROTEIN SMC01586.  
 GN SMC01586.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_Taxid=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21368234; PubMed=11474104;  
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,  
 RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,  
 RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,  
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,  
 RA Gloux S., Godrie T., Goffeau A., Golding B., Guzy J., Gurjal M.,  
 RA Hernandez-Lucas I., Hong A., Huitzer L., Hyman R.W., Jones T., Kahn D.,  
 RA Kahn M.L., Kalman S., Keating D.H., Kist E., Komp C., Lelaure V.,  
 RA Mesny D., Palm C., Beck M.C., Pohl T.M., Portetelle D., Punnelle B.,  
 RA Ramsperger U., Surzycki R., Thebaud P., Vandenbol M.,  
 RA Vorhoeffer F.J., Weidner S., Wells D.H., Wong K., Yen K.-C., Batut J.,  
 RT "The composite genome of the legume symbiont *Sinorhizobium meliloti*.";  
 RL Science 293:668-672(2001).

DR EMBL: AL591790; CAC46910.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 141 AA; 15677 MW; D3B4CB1BIC93F879 CRC64;

Query Match 34.9%; Score 49.5; DB 16; Length 141;  
 Best Local Similarity 52.9%; Pred. No. 9.7;  
 Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 4 PWT-AMYOHDLRAVRRF 19  
 ||| |||::: |||  
 Db 105 PWT-AMYRYCGSRVRSF 121

RESULT 8  
 ID 076609 PRELIMINARY; PRT; 417 AA.  
 AC 076609;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHEITICAL 47.0 KDA PROTEIN.  
 GN T07H3.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Beck C., O'Brien D., Kramer J.;  
 RT "The sequence of C. elegans cosmid T07H3.";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
 DR EMBL: AF077540; AAC26308.1; -  
 DR HSP: P13727; IHHU.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR001304; Lectin\_c.  
 DR Pfam: PF00431; CUB; 1.  
 DR Pfam: PF00059; lectin\_c; 2.  
 DR SMART: SM00034; CLECT; 2.  
 DR SMART: SM00042; CUB; 1.  
 DR PROSITE: PS01180; CUB; 1.  
 DR PROSITE: PS00615; C\_TYPE\_LECTIN\_1; UNKNOWN\_1.  
 DR PROSITE: PS50041; C\_TYPE\_LECTIN\_2; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 417 AA; 46974 MW; 090ABHFA20ABE5H CRC64;

Query Match 34.5%; Score 49; DB 5; Length 417;  
 Best Local Similarity 39.1%; Pred. No. 35;  
 Matches 9; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 2 RVPWTAWYOHDLRAVRRVHRRP 24  
 ||| |||::: |||  
 Db 256 ROPWGLWYTGNDYDRKFFCKRP 278

RESULT 9

Q94C00  
 ID Q94C00 PRELIMINARY; PRT; 431 AA.  
 AC Q94C00;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE P0660F12.30 PROTEIN.  
 GN P0660F12.30.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_Taxid=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0660F12.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003687; BAB3864.1; -  
 SQ SEQUENCE 431 AA; 46474 MW; 828E4E2554C2B8DB CRC64;

Query Match 34.5%; Score 49; DB 10; Length 431;  
 Best Local Similarity 60.0%; Pred. No. 36;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 AMYOHDLRAVRRFVH 21  
 ||| |||::: |||  
 Db 100 AMQPPDLRRFARFGH 114

RESULT 10  
 ID Q9AFX4 PRELIMINARY; PRT; 104 AA.  
 AC Q9AFX4;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ORF, HYPOTHETICAL.  
 GN S0038.  
 OS Shigella flexneri.  
 OG Plasmid virulence plasmid pWR501.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Shigella.  
 OX NCBI\_Taxid=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21189246; PubMed=11292750;  
 RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,  
 RA Blatner F.R.;  
 RT "Complete DNA sequence and analysis of the large virulence plasmid of  
 RT Shigella flexneri.";  
 RL Infect. Immun. 69:3271-3285(2001).  
 DR EMBL: AF348706; AAK18349.1; -  
 KW Plasmid.  
 SQ SEQUENCE 104 AA; 11767 MW; 7F0C625CA72ED4FF CRC64;

Query Match 34.2%; Score 48.5; DB 2; Length 104;  
 Best Local Similarity 35.0%; Pred. No. 9.9;  
 Matches 14; Conservative 2; Mismatches 7; Indels 17; Gaps 2;

QY 1 SRVPW-----TAWYOHDLRAVRRVHRRP 24  
 ||| |||::: |||  
 Db 22 SRVPWKTSTKSSAPSLKSTCTDW-KHYPLLROSHRRP 60

RESULT 11  
 ID Q30067 PRELIMINARY; PRT; 434 AA.  
 AC Q30067;



DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE HYPOHETICAL 50.6 KDA PROTEIN.  
 GN AFO170.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 NCBI\_TaxID=2234;  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Arlisch P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 DR EMBL: AE001094; AAB91065.1; -.  
 DR TIGR: AF0170; -.  
 DR InterPro: IPR000014; PAS.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 434 AA; 50640 MW; 777F8D5059347D2E CRC64;

Query Match 34.2%; Score 48.5; DB 17; Length 434;  
 Best Local Similarity 42.9%; Pred. No. 43;  
 Matches 9; Conservative 5; Mismatches 6; Indels 1; Gaps 1;  
 OY 3 VPWTAWYOHDLRAYRFRVHR 22  
 DB 162 IPNVLMFKHDEIRALRVLRH 182

RESULT 12  
 ID 09K0S2 PRELIMINARY; PRT; 506 AA.  
 AC 09K0S2;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE HYPOHETICAL PROTEIN NMB0506.  
 GN NMB0506.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OC NCBI\_TaxID=491;  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Clecko A., Parsey D.S., Blair E., Ciltone H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Masignani V., Pizaa M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 RT MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL: AE002407; AAF40938.1; -.  
 DR TIGR: NMB0506; -.  
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 506 AA; 52695 MW; A45ECDEA04046E5C CRC64;  
 Query Match 34.2%; Score 48.5; DB 16; Length 506;  
 Best Local Similarity 38.1%; Pred. No. 50;  
 Matches 8; Conservative 7; Mismatches 3; Indels 3; Gaps 1;  
 OY 2 RVPTAWYOHDLRAYRFRVHR 22  
 DB 342 REPWT---RHDVOTRYOYNO 359

RESULT 13  
 ID 09JPE1 PRELIMINARY; PRT; 833 AA.  
 AC 09JPE1;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)  
 DE HYPOHETICAL 88.4 KDA PROTEIN.  
 GN RTW4.  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FAM18;  
 RX MEDLINE=20187481; PubMed=10722605;  
 RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,  
 RA Tinsley C.R.;  
 RT "Molecular and biological analysis of eight genetic islands that  
 RT distinguish neisseria meningitidis from the closely related pathogen  
 RT neisseria gonorrhoeae.";  
 RL Infect. Immun. 68:2082-2095(2000).  
 DR EMBL: AJ391284; CAB72083.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 833 AA; 88396 MW; B967C8422CD71A12 CRC64;

Query Match 34.2%; Score 48.5; DB 2; Length 833;  
 Best Local Similarity 38.1%; Pred. No. 83;  
 Matches 8; Conservative 7; Mismatches 3; Indels 3; Gaps 1;  
 OY 2 RVPTAWYOHDLRAYRFRVHR 22  
 DB 669 REPWT---RHDVOTRYOYNO 686

RESULT 14  
 ID 09X4D5 PRELIMINARY; PRT; 269 AA.  
 AC 09X4D5;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, last annotation update)  
 DE LICD2 (LICD2 PROTEIN).  
 GN LICD2 OR SPI274.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6X;  
 RX MEDLINE=99217023; PubMed=10200966;  
 RA Zhang J.R., Idanpaan-Heikkila I., Fischer W., Tuomanen E.I.;  
 RT "Pneumococcal licd2 gene is involved in phosphorylcholine  
 RT metabolism.";  
 RL Mol. Microbiol. 31:1477-1488(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;

RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Kfoury H., Wolf A.M., Uterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldlyum T.V., Anginoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter C.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 RT pneumoniae.";  
 RL Science 293:498-506(2001).  
 DR EMBL; AF106539; AAD37094.1; -;  
 DR EMBL; AE007426; AAK75378.1; -;  
 DR TIGR; SP1274; -;  
 KW Complete proteome.  
 SQ SEQUENCE 269 AA; 32100 MW; 31A152DFAA480A10 CRC64;

Query Match 33.8%; Score 48; DB 16; Length 269;  
 Best Local Similarity 54.5%; Pred. No. 31;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 8 WYOHDLRAYRR 18  
 |||::|||:  
 DB 258 WYSHSIKAYRK 268

RESULT 15  
 053514  
 ID 053514 PRELIMINARY; PRT; 295 AA.  
 AC 053514  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE MEMBRANE PROTEIN.  
 GN RV2180C OR MTW021.13C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_taxid=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 DR EMBL; AL021957; CAAT7484.1; -;  
 DR TuberculList; RV2180C; -;  
 KW Complete proteome.  
 SQ SEQUENCE 295 AA; 33213 MW; 124501E68D4739CF CRC64;

Query Match 33.8%; Score 48; DB 16; Length 295;  
 Best Local Similarity 26.3%; Pred. No. 34;  
 Matches 10; Conservative 4; Mismatches 2; Indels 22; Gaps 1;

OY 8 WYOHDL-----RAYRRFYHRR 23  
 ||||:||||:|||||  
 DB 6 WLOHDIYDRGRLLPCLCVAFVLPFLVTRSFVFTIHR 43



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:49:01 ; Search time 73.67 seconds  
(without alignments)  
12.062 Million cell updates/sec

Title: US-09-613-092a-6\_COPY\_3\_10  
Perfect score: 46  
Sequence: 1 YOHDLRAY 8

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_032802:\*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	100.0	15	20	AAV30352
2	35	76.1	187	18	AAW20436
3	35	76.1	432	18	AAW20733
4	35	76.1	447	22	AAU36020
5	34	73.9	316	20	AAV42795
6	34	73.9	1136	22	ABG14341
7	34	73.9	1136	22	ABG15880
8	34	73.9	1249	22	ABG26042
9	33	71.7	314	18	AAW11801
10	33	71.7	316	19	AAW69546
11	33	71.7	398	18	AAW17998

12	33	71.7	400	19	AAW69545
13	33	71.7	451	21	AAW10565
14	33	71.7	466	21	AAW10564
15	33	71.7	468	21	AAW32000
16	33	71.7	480	21	AAW31999
17	33	71.7	488	21	AAW10563
18	33	71.7	490	21	AAW31998
19	33	71.7	1563	22	ABW58432
20	32	69.6	89	22	AAW42343
21	32	69.6	107	22	ABW08859
22	32	69.6	137	22	AAW33553
23	32	69.6	138	22	AAW36056
24	32	69.6	354	22	AAU38322
25	32	69.6	356	22	AAU34471
26	32	69.6	370	22	AAU44615
27	32	69.6	464	22	ABG14467
28	32	69.6	512	22	ABW69433
29	32	69.6	604	22	ABW71943
30	32	69.6	676	22	ABW08862
31	32	69.6	708	21	AAW97550
32	32	69.6	1788	21	AAW85575
33	32	69.6	2396	22	ABW64047
34	31	67.4	28	17	AAW91380
35	31	67.4	86	21	AAW00043
36	31	67.4	239	22	ABW29619
37	31	67.4	295	22	ABW29618
38	31	67.4	380	20	AAW32053
39	31	67.4	432	20	AAW88236
40	31	67.4	441	22	ABW92783
41	31	67.4	442	22	ABW71447
42	31	67.4	449	21	AAW12572
43	31	67.4	533	10	AAW3091
44	31	67.4	554	22	ABW19731
45	31	67.4	554	22	ABW19785

## ALIGNMENTS

RESULT	1	ALIGNMENTS
AAV30352	standard; Peptide: 15 AA.	
AAV30352:		
09-NOV-1999	(first entry)	
DE	Epitope derived from pneumococcal surface adhesion A protein.	
XX	Pneumococcal surface adhesion A protein; Psat: monoclonal antibody;	
KW	vaccine; Streptococcus pneumoniae infection.	
XX		
OS	Streptococcus pneumoniae.	
XX		
PN	WO945121-A1.	
XX		
PD	10-SEP-1999.	
XX		
PF	26-FEB-1999; 99WO-US04326.	
XX		
PR	02-MAR-1998; 98US-0076565.	
XX		
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
PI	Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ,	
PI	Zeller JL;	
XX	WPI; 1999-540849/45.	
DR		
XX		
PT	New peptides corresponding to Streptococcus pneumoniae Psat, used	
PT	for treating or preventing Streptococcus pneumoniae infection in a	
XX	subject	

Dirofilaria immiti  
Aradipops thalia  
Aradipops thalia  
Aradipops thalia  
Aradipops thalia  
Aradipops thalia  
Aradipops thalia  
Drosophila melanog  
Protonibacterium  
Novel human diagno  
Klebsiella pneumon  
Klebsiella pneumon  
Salmonella typhi c  
E. coli cellular p  
Protonibacterium  
Novel human diagno  
Drosophila melanog  
Drosophila melanog  
Novel human diagno  
Novel human diagno  
Mouse PAMP protein  
D. melanogaster BA  
Drosophila melanog  
H2-Kappa-B binding  
Human secreted pro  
Novel human diagno  
Novel human diagno  
Bovine pregnancy a  
Mouse prothrombina  
Human protein sequ  
Drosophila melanog  
NDO related comple  
Vire2 gene product  
Novel human diagno  
Novel human diagno

PS Claim 6; Page 43; 58pp; English.

XX  
CC AAW30351-54 represent immunogenic peptides which are derived from  
CC a pneumococcal surface adhesion A protein (PsaA). The specific  
CC descriptors monoclonal antibodies which bind epitopes of the PsaA protein  
CC (e.g present sequence). The peptides can be used in vaccines to prevent  
CC Streptococcus pneumoniae infections. The antibodies of the invention  
CC can also be used to detect S. pneumoniae in a sample or individual.  
XX

SO Sequence 15 AA;

OY 1 YOHDLRAY 8  
| | | | |  
Db 3 yghdIray 10

RESULT 2  
AAW20436  
ID AAW20436 standard; Protein: 187 AA.  
XX  
AC AAW20436;  
XX  
DT 14-JUL-1997 (first entry)  
XX  
DE H. pylori protein.  
XX  
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;  
KW outer membrane; cell envelope; transporter.  
XX  
OS Helicobacter pylori.  
XX  
FN W09640893-A1.  
XX  
PD 19-DEC-1996.  
XX  
PF 06-JUN-1996; 96WO-US09122.  
PR 01-APR-1996; 96US-0630405.  
PR 07-JUN-1995; 95US-0487032.  
XX  
PA (ASTRA ) ASTRA AB.  
XX  
PI Berglindh OT, Smith D, Meligaerd BL,  
XX  
DR WPI: 1997-053306/05.  
XX  
N-PSDB: AAT67609.  
XX  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
XX  
XX  
XX Disclosure: Pages 613; 1481pp; English.

XX  
CC The present sequence is a Helicobacter pylori protein of unknown  
CC function. The protein may be used in a vaccine to prevent or treat  
CC H. pylori infection or to identify H. pylori polypeptide binding  
CC compounds, useful as potential H. pylori life cycle activators or  
CC inhibitors. The genomic sequence of H. pylori (ATCC 55879) was  
CC determined from overlapping contigs generated by mechanically shearing  
CC the bacterial DNA. The sequences were analysed for ORF of at least 180  
CC nucleotides, and the predicted coding regions defined by computer  
CC evaluation. To identify likely H. pylori antigens for vaccine  
CC development, the amino acid sequences predicted from various ORF were  
CC analysed for significant homology to other known or exported membrane  
CC proteins. Having identified and determined the sequences of interest,  
CC particular regions can be isolated from H. pylori by PCR amplification

```

xx      for recombinant polypeptide production, e.g. in E. coli hosts.
cc
SQ      Sequence      187 AA;

Query Match      76.1%; Score 35; DB 18; Length 187;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 YOHDR 6
        |||||
Db      103 yphdir 108

RESULT      3
ID      AAM20733
        AAM20733 standard; protein; 432 AA.
XX
AC      AAM20733;
XX
DT      16-JUL-1997 (first entry)
XX
DE      H. pylori cell envelope protein, 06cpl1722orf15.
XX
KW      Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW      binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW      duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS      Helicobacter pylori.
XX
PN      W09640893-A1.
XX
PD      19-DEC-1996.
XX
PF      06-JUN-1996; 96WO-US09122.
XX
PR      01-APR-1996; 96US-0630405.
XX
PR      07-JUN-1995; 95US-0487032.
XX
PA      (ASTR ) ASTRA AB.
XX
PI      Berglindh OT, Smith D, Mellgaard BL;
XX
DR      WPI: 1997-052306/05.
DR      N-FSDB; AAT67986.
XX
PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
XX
PS      Claim 56; Page 1148-1149; 1481pp; English.
XX
CC      The present sequence is a Helicobacter pylori cell envelope protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed by computer evaluation. To
CC      and the predicted coding regions defined by computer evaluation. To
CC      identify likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide
CC      production, e.g. in E. coli hosts.
XX
SQ      Sequence      432 AA;

Query Match      76.1%; Score 35; DB 18; Length 432;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 YOHDLR 6  
| | | | |  
DB 239 yghdlr 244

RESULT 4  
AAU36020  
ID AAU36020 standard; Protein; 447 AA.

AC AAU36020;

DE 14-FEB-2002 (first entry)

DE Helicobacter pylori cellular proliferation protein #333.

KW Antisense; prokaryotic cellular proliferation protein;  
antibiotic; antibacterial; drug design.

OS Helicobacter pylori.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR WPI: 2001-611495/70.

DR N-PSDB: AAS53879.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 11613; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 447 AA;

Query Match 76.1%; Score 35; DB 22; Length 447;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YOHDLR 6  
| | | | |  
DB 238 yghdlr 243

RESULT 5  
AAV42795  
ID AAV42795 standard; Protein; 316 AA.

AC AAV42795;

DE 05-JAN-2000 (first entry)

DE Streptococcus pyogenes SAG-A processing protease.

KW Streptococcus; SAG-A; tissue necrosis; antibacterial; streptolysin S;

KW activity; oxygen-stable; non-immunogenic; cytotoxin; beta-haemolysis;

KW disease; strep throat; cellulitis; scarlet fever; impetigo;

KW rheumatic fever; acute glomerular nephritis; endocarditis;

KW necrotising fasciitis; brain abscess; meningitis; osteomyelitis;

KW pharyngitis; pneumonia; rheumatic carditis; toxic shock; protease.

OS Streptococcus pyogenes.

PN WO9949049-A1.

PD 30-SEP-1999.

PF 18-MAR-1999; 99WO-CA00240.

PR 20-MAR-1998; 98US-0078713.

PR (MOUN) MOUNT SINAI HOSPITAL.

PI De Azavedo J, Bast D, Borgia S, Betschel S, Low D;

PI WPI: 1999-591100/50.

DR N-PSDB: AA230815.

XX New isolated Streptococcus polypeptides, used to develop products for

PT treating, e.g. streptococcal infections, microbial infections in plants

PS and animals and cancers and as preservatives -

XX Disclosure; Page 92; 99pp; English.

CC This sequence represents a SAG-A processing protease. This

CC is an accessory protein used in the synthesis of a novel

CC Streptococcus pyogenes protein, SAG-A (AAV42792), which

CC is associated with streptolysin S (SLS) activity. Strains of

CC S. pyogenes which were SLS-deficient and had significantly reduced

CC virulence were found to have a single transposon (Tn) 916 insertion in

CC the promoter region of this gene. The SAG-A protein has subsequently

CC been shown to play a direct role in tissue necrosis. SLS is produced by

CC virtually all strains of group A streptococci (GAS) and has a direct

CC cytopathic effect on a broad range of cell types. It is an

CC oxygen-stable, non-immunogenic, cytotoxin which causes beta-haemolysis.

CC Streptococcal infections cause disorders such as endocarditis,

CC cellulitis, brain abscesses, glomerulonephritis, pneumonia, meningitis,

CC osteomyelitis, pharyngitis, rheumatic fever, pneumonia, strep throat,

CC scarlet fever, impetigo, necrotising fasciitis, rheumatic carditis, and

CC toxic shock. SAG-A proteins and peptides can be used to identify SAG-A

CC agonists or antagonists, preferably bacteriostatic or bactericidal

CC agonists and antagonists. Inhibitors or antagonists can be used for

CC treating the disorders caused by infection with streptococci. The SAG-A

CC peptides may be useful in both the pharmaceutical and food industries.

CC They may exhibit antibacterial activity against a wide variety of

CC gram-negative and gram-positive bacteria and may be used as a food

preservative, or as an antibacterial agent for medical or agricultural use. They may function synergistically with conventional therapeutic agents such as antibiotics and anticancer treatments, and they may be used as adjuvants. They may also be used to selectively lyse cells using a chimeric toxin comprising a cytolytic polypeptide operatively linked to a targeting agent. SAG-A peptides can also be used for detection and diagnosis and in vaccines for preventing infections.

... SQ Sequence 316 AA;

Query Match	73.98;	Score 34;	DB 20;	Length 316;
Best Local Similarity	75.08;	Pred. No. 54;		
Matches 6; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	1	YOHDLRAY	8
	1	1	1
	1	1	1
	1	1	1
Db	204	YQHAIrcy	211

RESULT	6
ABG14341	
ID	ABG14341 standard; Protein; 1136 AA.

AC ABG14341;

DT 18-FEB-2002 (first entry)

Novel human diagnostic protein #14332.

Human: chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

XX	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
----	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

XX	/	19
XX	/	19

XX

DR N-PSDB; AAS78528.

PT New isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity -

Claim 20; SEQ ID No 44700; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.

CC diagnostic amino acid sequences of the invention.

CC specification, but was obtained in electronic format directly from WIPO

CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 1136 AA;

Query Match	73.9%;	Score 34;	DB 22;	Length 1136;
Best Local Similarity	62.5%;	Pred. No. 2.2e+02;		
Matches	5;	Conservative	2;	Mismatches 1;
			Indels	0;
			Gaps	0;

Qy	1	YQHDLRAY	8
	:	:	:
	:	:	:
	:	:	:
Db	615	yrhndlkily	622

RESULT	7
ABG15880	
ID	ABG15880 standard; Protein; 1136 AA.

AC ABG15880

DT 18-FEB-2002 (first entry)

Novel human diagnostic protein #15871.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

[illegible]

XX  
XX

noho sapientis.

XX  
XX

0001700ZOM  
N1

FD 11 OCT 2001  
XX

XX 20 JAN 2001, 2001MO 050000Z

23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess

f1  
 XX  
 universally

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and



CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

CC Sequence 1136 AA;

Query Match 73.9%; Score 34; DB 22; Length 1136;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDRLRAY 8  
1:||||:1  
Db 615 ythdikiy 622

#### RESULT 8

ABG26042  
ID ABG26042 standard; Protein; 1249 AA.

AC ABG26042;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26033.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS90229.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 20; SEQ ID NO 56401; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

CC Sequence 1249 AA;

Query Match 73.9%; Score 34; DB 22; Length 1249;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDRLRAY 8  
1:||||:1  
Db 615 ythdikiy 622

#### RESULT 9

AAW11801  
ID AAW11801 standard; Protein; 314 AA.

AC AAW11801;

DT 06-MAY-1997 (first entry)

DE Dirofilaria immitis L3 cysteine protease PDICP314.

KM Cysteine protease; filarial nematode; helminth; vaccine; heartworm;  
KM Onchocerca; ndicP942; PDICP314.

OS Dirofilaria immitis.

PN WO9640884-A1.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US09848.

PR 07-JUN-1995; 95US-0486036.

PA (HESK-) HESKA CORP.

PA (COLS-) UNIV COLORADO STATE RES FOUND.

PI Frank GR, Grieve RB, Tripp CA, Wisniewski N;

DR WPI: 1997-099931/09.

DR N-PSDB; AAT59480.

PT Filarial nematode larval nucleic acid - capable of hybridizing with  
PT dirofilaria immitis or Onchocerca volvulus L3 cysteine protease, to  
PT protect against parasitic helminth diseases

PS Claim 16; Page 89-90; 115pp; English.

CC Cysteine protease (CP) polypeptide PDICP314 (AAW11801) comprises the  
CC mature CP protein encoded by nucleic acid ndicP942 (AAT59478),  
CC derived from a cDNA library of Dirofilaria immitis L3 larvae.  
CC Novel filarial nematode CP polypeptides (AAW11798-802) are capable of  
CC eliciting an immune response against native helminth CPs. They can  
CC be used to identify CPs, capable of inhibiting the CP activity of  
CC a parasitic helminth and to raise anti-CP antibodies. The  
CC polypeptide (expressed e.g. by transformed host cells), inhibitor  
CC or antibody can be used in therapeutic compns. to protect an  
CC animal from a disease caused by a parasitic helminth, such as  
CC dirofilaria or Onchocerca.

CC Sequence 314 AA;

Query Match 71.7%; Score 33; DB 18; Length 314;  
Best Local Similarity 62.5%; Pred. No. 86;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
 |||| : |  
 Db 239 yghdfkfy 246

RESULT 10  
 AAM69546  
 ID AAM69546 standard; Protein: 316 AA.

AC AAM69546;  
 XX  
 DT 09-OCN-1998 (first entry)  
 XX

DE Dirofilaria immitis L3 larval cysteine protease #2.

XX Onchocerca volvulus; Dirofilaria immitis; L3 larval cysteine protease;  
 KW filarid nematode; parasitic helminth; vaccine.  
 XX

OS Dirofilaria immitis.

PN US5792624-A.

PD 11-AUG-1998.

PF 07-JUN-1995; 95US-0482282.

PR 07-JUN-1995; 95US-0482282.

PR 12-FEB-1991; 91US-0654226.

PR 12-NOV-1991; 91US-0792209.

PR 03-AUG-1993; 93US-0101283.

PR 16-NOV-1993; 93US-0153554.

XX (HESK-) HESKA CORP.

PA (COLS ) UNIV COLORADO STATE RES FOUND.

PI Frank GR, Grieve RB, Richer JK, Tripp CA, Wisniewski N;

DR N-PSDB; AAV40249.

DR N-PSDB; AAV40249.

XX Nematode larval protease proteins - useful for vaccination, etc.

PS Example 1; Column 35-38; 22pp; English.

XX The present sequence represents an L3 larval protease protein from

CC Dirofilaria immitis. An embodiment of the present invention is an

CC isolated filarid nematode nucleic acid molecule that hybridises, under

CC stringent hybridisation conditions, with a dirofilaria immitis L3 larval

CC cysteine protease gene and/or an Onchocerca volvulus L3 larval cysteine

CC protease gene. A filarid nematode cysteine protease protein of the

CC present invention preferably has cysteine protease activity and/or

CC comprises a protein that, when administered to an animal, is capable of

CC eliciting an immune response against a natural helminth cysteine protease

CC protein. This sequence can be used in a therapeutic composition

CC capable of protecting an animal from disease caused by a parasitic

CC helminth.  
 XX

SQ Sequence 316 AA;

Query Match 71.7%; Score 33; DB 19; Length 316;

Best Local Similarity 62.5%; Pred. No. 86;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
 |||| : |  
 Db 241 yghdfkfy 248

RESULT 11

AAM11798  
 ID AAM11798 standard; Protein: 398 AA.

AC AAM11798;

XX 06-MAY-1997 (first entry)

XX Dirofilaria immitis L3 cysteine protease PDICP398.

XX Cysteine protease; filarid nematode; helminth; vaccine; heartworm;

XX Onchocerca; ndICP1298; PDICP398.

OS Dirofilaria immitis.

XX Key Location/Qualifiers

FT Protein 85..398

FT /label= Mat\_protein

FT /note= "polypeptide PDICP314"

PN WO9640884-A1.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US09848.

PR 07-JUN-1995; 95US-0486036.

XX (HESK-) HESKA CORP.

PA (COLS ) UNIV COLORADO STATE RES FOUND.

PI Frank GR, Grieve RB, Tripp CA, Wisniewski N;

DR N-PSDB; AAT59477.

DR N-PSDB; AAT59477.

XX Filarid nematode larval nucleic acid - capable of hybridising with

PT Dirofilaria immitis or Onchocerca volvulus L3 cysteine protease, to

PT protect against parasitic helminth diseases

PS Claim 16; Page 86-87; 115pp; English.

XX Cysteine protease (CP) polypeptide PDICP398 (AAM11798) is encoded by

CC nucleic acid ndICP1298 (AAT59477), ctd. from a cDNA library of

CC Dirofilaria immitis L3 larvae. Novel filarid nematode CP

CC polypeptides (AAM11798-802) are capable of eliciting an immune

CC response against native helminth CPs. They can be used to identify

CC cpds, capable of inhibiting the CP activity of a parasitic

CC helminth and to raise anti-CP antibodies. The polypeptide

CC (expressed e.g. by transformed host cells), inhibitor or antibody

CC can be used in therapeutic compns. to protect an animal from a

CC disease caused by a parasitic helminth, such as Dirofilaria or

CC Onchocerca.  
 XX

SQ Sequence 398 AA;

Query Match 71.7%; Score 33; DB 18; Length 398;

Best Local Similarity 62.5%; Pred. No. 1.1e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
 |||| : |  
 Db 323 yghdfkfy 330

RESULT 12

AAM69545  
 ID AAM69545 standard; Protein: 400 AA.

AC AAM69545;

XX 09-OCN-1998 (first entry)

XX Dirofilaria immitis L3 larval cysteine protease #1.

DE Onchocerca volvulus; Dirofilaria immitis; L3 larval cysteine protease;  
 KW

KW filarid nematode; parasitic helminth; vaccine.  
XX  
OS Dirofilaria immitis.  
XX  
PN US5792624-A.  
XX  
PD 11-AUG-1998.  
XX  
PF 07-JUN-1995; 95US-0482282.  
XX  
PR 07-JUN-1995; 95US-0482282.  
PR 12-FEB-1991; 91US-0654226.  
PR 12-NOV-1991; 91US-0792209.  
PR 03-AUG-1993; 93US-0101283.  
PR 16-NOV-1993; 93US-0153554.  
XX  
PA (HESK-) HESKA CORP.  
PA (COLS ) UNITV COLORADO STATE RES FOUND.  
XX  
PI Frank GR, Griewe RB, Richer JK, Tripp CA, Wisniewski N;  
XX  
DR WPI; 1998-456128/39.  
DR N-PSDB; AAV40248.  
XX  
PT Nematode larval protease proteins - useful for vaccination, etc.  
PS  
XX Example 1; Column 31-34; 22pp; English.  
XX  
CC The present sequence represents an L3 larval protease protein from  
CC Dirofilaria immitis. An embodiment of the present invention is an  
CC isolated filarid nematode nucleic acid molecule that hybridises, under  
CC stringent hybridisation conditions, with a Dirofilaria immitis L3 larval  
CC cysteine protease gene and/or an Onchocerca volvulus L3 larval cysteine  
CC protease gene. A filarid nematode cysteine protease protein of the  
CC present invention preferably has cysteine protease activity and/or  
CC comprises a protein that, when administered to an animal, is capable of  
CC eliciting an immune response against a natural helminth cysteine protease  
CC protein. This sequence can be used in a therapeutic composition  
CC capable of protecting an animal from disease caused by a parasitic  
CC helminth.  
XX  
SO Sequence 400 AA;

Query Match 71.7%; Score 33; DB 19; Length 400;  
Best Local Similarity 62.5%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDIRAY 8  
Db 325 yghdfkfy 332

RESULT 13  
AAG10565  
ID AAG10565 standard; Protein; 451 AA.  
XX  
AC AAG10565;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8937.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX

PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139465.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.

09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 18-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149910.  
PR 25-AUG-1999; 99US-0150586.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159684.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 71.7%; Score 33; DB 21; Length 451;  
Best Local Similarity 62.5%; Pred. No. 1; 3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDRAY 8  
Db 78 ydhalesy 85

RESULT 14  
ID AAG10564 standard; Protein; 466 AA.  
AC AAG10564;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8936.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 9905-0126785.  
PR 01-APR-1999; 9905-0127462.  
PR 06-APR-1999; 9905-0128234.  
PR 08-APR-1999; 9905-0128714.  
PR 16-APR-1999; 9905-0129845.  
PR 19-APR-1999; 9905-0130077.  
PR 21-APR-1999; 9905-0130449.  
PR 23-APR-1999; 9905-0130510.  
PR 28-APR-1999; 9905-0130891.  
PR 30-APR-1999; 9905-0131449.  
PR 30-APR-1999; 9905-0132048.  
PR 30-APR-1999; 9905-0132407.  
PR 04-MAY-1999; 9905-0132484.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 06-MAY-1999; 9905-0132487.  
PR 07-MAY-1999; 9905-0132863.  
PR 11-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134321.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147316.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.

PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 71.7%; Score 33; DB 21; Length 466;  
Best Local Similarity 62.5%; Pred. No. 1,3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
| | | | : |  
Db 93 ydhdlresy 100

RESULT 15  
AAG32000  
ID AAG32000 standard; Protein; 468 AA.

XX AAG32000;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 38524.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX PD 06-SEP-2000.

XX PE 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 71.7%; Score 33; DB 21; Length 468;  
Best Local Similarity 62.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
Db 92 ydhdlresy 99

Search completed: August 15, 2002, 11:49:02  
Job time: 358 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:49:35 ; Search time 27.13 Seconds  
(without alignments)  
7.203 Million cell updates/sec

Title: US-09-613-092a-6\_COPY\_3\_10  
Perfect score: 46  
Sequence: 1 YOHDLRAY 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Dackfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	73.9	67	2	US-08-248-839C-100
2	33	71.7	213	5	PCT-US96-09848-21
3	33	71.7	314	1	US-08-486-036A-4
4	33	71.7	314	5	PCT-US96-09848-4
5	33	71.7	316	1	US-08-482-282B-4
6	33	71.7	356	5	PCT-US96-09848-24
7	33	71.7	398	1	US-08-486-036A-2
8	33	71.7	398	1	PCT-US96-09848-2
9	33	71.7	400	1	US-08-482-282B-2
10	31	67.4	887	1	US-08-066-167-3
11	31	67.4	911	1	US-08-487-890A-95
12	31	67.4	911	2	US-08-478-435-95
13	31	67.4	911	2	US-08-337-483-95
14	31	67.4	911	2	US-08-478-373-95
15	31	67.4	911	3	US-08-474-671-95
16	31	67.4	911	3	US-08-483-577A-95
17	31	67.4	911	3	US-08-448-194-6
18	31	67.4	911	3	US-08-613-009A-17
19	31	67.4	911	4	US-08-897-438-95
20	31	67.4	911	4	US-08-867-921-6
21	31	67.4	911	4	US-08-637-654-95
22	30	65.2	243	4	US-08-944-483-70
23	30	65.2	567	2	US-08-504-459-2
24	30	65.2	567	2	US-08-504-459-6
25	30	65.2	621	4	US-09-059-522-1
26	30	65.2	621	4	US-09-382-027-1
27	30	65.2	686	2	US-08-849-480A-4

28	30	65.2	770	1	US-08-369-796-12	Sequence 12, Appl
29	30	65.2	770	1	US-08-416-581B-1	Sequence 1, Appl
30	30	65.2	770	1	US-08-416-581B-5	Sequence 5, Appl
31	30	65.2	770	1	US-08-416-581B-9	Sequence 9, Appl
32	30	65.2	770	2	US-08-852-091-12	Sequence 12, Appl
33	30	65.2	770	2	US-08-820-754-12	Sequence 12, Appl
34	30	65.2	770	3	US-08-956-652-12	Sequence 12, Appl
35	30	65.2	770	3	US-08-956-869-12	Sequence 12, Appl
36	30	65.2	770	3	US-09-012-710-8	Sequence 8, Appl
37	30	65.2	770	3	US-08-948-547-12	Sequence 12, Appl
38	30	65.2	770	4	US-09-087-465-6	Sequence 6, Appl
39	30	65.2	770	4	US-09-364-970-3	Sequence 3, Appl
40	30	65.2	770	4	US-09-364-970-5	Sequence 5, Appl
41	30	65.2	770	4	US-09-556-273-8	Sequence 8, Appl
42	30	65.2	770	4	US-08-956-653A-12	Sequence 12, Appl
43	30	65.2	770	5	PCT-US95-17025-12	Sequence 12, Appl
44	30	65.2	771	1	US-08-276-099A-14	Sequence 14, Appl
45	30	65.2	771	1	US-08-781-890-14	Sequence 14, Appl

#### ALIGNMENTS

RESULT 1  
US-08-248-839C-100  
Sequence 100, Application US/08248839C  
Patent No. 5843702  
GENERAL INFORMATION:  
APPLICANT: McConnell, David  
APPLICANT: Devine, Kevin  
APPLICANT: O' Kane, Charles  
TITLE OF INVENTION: A Gene Expression System  
NUMBER OF SEQUENCES: 185  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 58437020 No. 5843702disk of No. 5843702th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PastSEO for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/248,839C  
FILING DATE: 25-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valera A.  
REGISTRATION NUMBER: 35,127  
REFERENCE/DOCKET NUMBER: 3614.214-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 100:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 67 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-248-839C-100

Query Match 73.9%; Score 34; DB 2; Length 67;  
Best Local Similarity 62.5%; Pred. No. 5.7;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
Db 27 YNHDLRSF 34

```
RESULT 2
PCT-US96-09848-21
: Sequence 21, Application PC/TUS9609848
: GENERAL INFORMATION:
: APPLICANT: Tripp, Cynthia A.
: APPLICANT: Wisniewski, Nancy
: TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
: TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/09848
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/486,036
: FILING DATE: 07-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2618-33-PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 213 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US96-09848-21

Query Match 71.7%: Score 33; DB 5; Length 213;
Best Local Similarity 62.5%: Pred. No. 28;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8
||| : |
Db 138 YOHDKFY 145

RESULT 3
US-08-486-036A-4
: Sequence 4, Application US/08486036A
: Patent No. 5795768
: GENERAL INFORMATION:
: APPLICANT: Tripp, Cynthia A.
: APPLICANT: Wisniewski, Nancy
: APPLICANT: Grieve, Robert B.
: APPLICANT: Frank, Glenn R.
: TITLE OF INVENTION: NOVEL FILARIID NEMATODE
: TITLE OF INVENTION: CYSTEINE PROTEASE PROTEINS, NUCLEIC ACID
: MOLECULES AND USES THEREOF
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, Suite 3500
: CITY: Denver
: STATE: Colorado
```

```
COUNTRY: U.S.A.
ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/486,036A
: FILING DATE: June 7, 1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2618-33
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 314 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-486-036A-4

Query Match 71.7%: Score 33; DB 1; Length 314;
Best Local Similarity 62.5%: Pred. No. 41;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8
||| : |
Db 239 YOHDKFY 246

RESULT 4
PCT-US96-09848-4
: Sequence 4, Application PC/TUS9609848
: GENERAL INFORMATION:
: APPLICANT: Tripp, Cynthia A.
: APPLICANT: Wisniewski, Nancy
: TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
: TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/09848
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/486,036
: FILING DATE: 07-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2618-33-PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
```

LENGTH: 314 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-09848-4

Query Match 71.7%; Score 33; DB 5; Length 314;  
Best Local Similarity 62.5%; Pred. No. 41;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
||||:|  
DB 239 YOHDFKFY 246

RESULT 5  
US-08-482-282B-4  
Sequence 4, Application US/08482282B  
Patent No. 5792624  
GENERAL INFORMATION:  
APPLICANT: Tripp, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Griewe, Robert B.  
APPLICANT: Frank, Glenn R.  
TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE  
TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carol Talkington Verser, Ph.D.  
ADDRESSEE: Haska Corporation  
STREET: 1825 Sharp Point Drive  
CITY: Fort Collins  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80525  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: Wordperfect for Windows, Version 7.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,282B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington  
REGISTRATION NUMBER: 37,459  
REFERENCE/DOCKET NUMBER: 2618-33-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (970) 493-7272  
TELEFAX: (970) 484-9505  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 316 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-282B-4

Query Match 71.7%; Score 33; DB 1; Length 316;  
Best Local Similarity 62.5%; Pred. No. 41;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
||||:|  
DB 241 YOHDFKFY 248

RESULT 6  
PCT-US96-09848-24

Sequence 24, Application PC/TUS9609848  
GENERAL INFORMATION:  
APPLICANT: Tripp, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE  
TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/09848  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,036  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-33-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 356 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1172  
PCT-US96-09848-24

Query Match 71.7%; Score 33; DB 5; Length 356;  
Best Local Similarity 62.5%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
||||:|  
DB 281 YOHDFKFY 288

RESULT 7  
US-08-486-036A-2  
Sequence 2, Application US/08486036A  
Patent No. 5795768  
GENERAL INFORMATION:  
APPLICANT: Tripp, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Griewe, Robert B.  
APPLICANT: Frank, Glenn R.  
TITLE OF INVENTION: NOVEL FILARIID NEMATODE  
TITLE OF INVENTION: CYSTEINE PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.

ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,036A  
FILING DATE: June 7, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-33  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 398 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-486-036A-2

Query Match 71.7%; Score 33; DB 1; Length 398;  
Best Local Similarity 62.5%; Pred. NO. 52;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YOHDLRAY 8  
||| : |  
Db 323 YOHDFKFY 330

RESULT 8  
PCT-US96-09848-2  
Sequence 2, Application PC/TUS9609848  
GENERAL INFORMATION:  
APPLICANT: Tripp, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE  
TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/09848  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,036  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-33-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 398 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-09848-2

Query Match 71.7%; Score 33; DB 5; Length 398;  
Best Local Similarity 62.5%; Pred. NO. 52;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YOHDLRAY 8  
||| : |  
Db 323 YOHDFKFY 330

RESULT 9  
US-08-482-282B-2  
Sequence 2, Application US/08482282B  
Patent No. 5792624  
GENERAL INFORMATION:  
APPLICANT: Tripp, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Grieve, Robert B.  
APPLICANT: Frank, Glenn R.  
TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE  
TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carol Talkington Verser, Ph.D.  
STREET: 1825 Sharp Point Drive  
CITY: Fort Collins  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80525  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: Wordperfect for Windows, Version 7.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,282B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington  
REGISTRATION NUMBER: 37,459  
REFERENCE/DOCKET NUMBER: 2618-33-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (970) 493-7272  
TELEFAX: (970) 484-9505  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-282B-2

Query Match 71.7%; Score 33; DB 1; Length 400;  
Best Local Similarity 62.5%; Pred. NO. 52;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YOHDLRAY 8  
||| : |  
Db 325 YOHDFKFY 332

RESULT 10  
US-08-066-167-3  
Sequence 3, Application US/08066167

```
? Patent No. 5618541
? GENERAL INFORMATION:
? APPLICANT: OQUENTIN-MILLET, Marie-Jose
? TITLE OF INVENTION: VACCINE AGAINST NEISSERIA MENINGITIDIS
? TITLE OF INVENTION: INFECTIONS
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Burns, Doane, Swecker & Mathis
? STREET: George Mason Bldg., Washington & Prince Sts.
? CITY: Alexandria
? STATE: Virginia
? COUNTRY: United States
? ZIP: 22313-1404
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/066,167
? FILING DATE: 02-JUN-1993
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: FR 91 12177
? FILING DATE: 03-OCT-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Crane-Feury, Sharon E
? REGISTRATION NUMBER: 36,113
? REFERENCE/DOCKET NUMBER: 016100-002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 836-6620
? TELEFAX: (703) 836-6620
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 887 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM: Neisseria meningitidis 2169 subunit rbp1.
? US-08-066-167-3

Query Match 67.4%; Score 31; DB 1; Length 887;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8
DB 493 YOHANRAY 500

RESULT 11
US-08-487-890A-95
? Sequence 95, Application US/08487890A
? Patent No. 5708149
? GENERAL INFORMATION:
? APPLICANT: Loosmore, Sheena
? APPLICANT: Harkness, Robin
? APPLICANT: Schryvers, Anthony
? APPLICANT: Chong, Pele
? APPLICANT: Gray-Owen, Scott
? APPLICANT: Yang, Yan-Ping
? APPLICANT: Murdin, Andrew
? APPLICANT: Klein, Michel
? TITLE OF INVENTION: Transferrin Receptor Genes
? NUMBER OF SEQUENCES: 147
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Sim & McBurney
? STREET: 6th Floor, 330 University Avenue
? CITY: Toronto
? STATE: Ontario
```

```
? COUNTRY: Canada
? ZIP: M5G 1R7
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/487,890A
? FILING DATE: 07-JUN-1993
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/175,116
? FILING DATE: 29-DEC-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/148,968
? FILING DATE: 08-NOV-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Stewart, Michael I
? REGISTRATION NUMBER: 24,973
? REFERENCE/DOCKET NUMBER: 1038-466 MIS:jfb
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (416) 595-1155
? TELEFAX: (416) 595-1163
? INFORMATION FOR SEQ ID NO: 95:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 911 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? US-08-487-890A-95
```

```
Query Match 67.4%; Score 31; DB 1; Length 911;
Best Local Similarity 75.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8
DB 517 YOHANRAY 524
```

```
RESULT 12
US-08-478-435-95
? Sequence 95, Application US/08478435
? Patent No. 5923323
? GENERAL INFORMATION:
? APPLICANT: Loosmore, Sheena
? APPLICANT: Harkness, Robin
? APPLICANT: Schryvers, Anthony
? APPLICANT: Chong, Pele
? APPLICANT: Gray-Owen, Scott
? APPLICANT: Yang, Yan-Ping
? APPLICANT: Murdin, Andrew
? APPLICANT: Klein, Michel
? TITLE OF INVENTION: Transferrin Receptor Genes
? NUMBER OF SEQUENCES: 147
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Sim & McBurney
? STREET: Suite 701, 330 University Avenue
? CITY: Toronto
? STATE: Ontario
? COUNTRY: Canada
? ZIP: M5G 1R7
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/478,435
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 435
```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
SEQUENCE CHARACTERISTICS:  
LENGTH: 911 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-478-435-95

Query Match 67.4%; Score 31; DB 2; Length 911;  
Best Local Similarity 75.0%; Pred. No. 2.8e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
DB 517 YOHANRAY 524

RESULT 13  
US-08-337-483-95  
Sequence 95, Application US/08337483  
Patent No. 5922562  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Mordin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 95:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 911 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-337-483-95

Query Match 67.4%; Score 31; DB 2; Length 911;  
Best Local Similarity 75.0%; Pred. No. 2.8e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
DB 517 YOHANRAY 524

RESULT 14  
US-08-478-373-95  
Sequence 95, Application US/08478373  
Patent No. 5922841  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Mordin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,373  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
SEQUENCE CHARACTERISTICS:  
LENGTH: 911 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-478-373-95

Query Match 67.4%: Score 31: DB 2: Length 911:  
Best Local Similarity 75.0%: Pred. No. 2.8e+02:  
Matches 6: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

Db 517 YOHANRAY 524

Search completed: August 15, 2002, 11:49:35  
Job time: 276 sec

Oy 1 YOHDLRAY 8  
Db 517 YOHANRAY 524

RESULT 15

US-08-474-671-95  
Sequence 95, Application US/08474671  
Patent No. 6008326

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena

APPLICANT: Harkness, Robin

APPLICANT: Schryvers, Anthony

APPLICANT: Chong, Pele

APPLICANT: Gray-Owen, Scott

APPLICANT: Yang, Yan-ping

APPLICANT: Mardin, Andrew

APPLICANT: Kieln, Michel

TITLE OF INVENTION: Transferrin Receptor Genes

NUMBER OF SEQUENCES: 147

CORRESPONDENCE ADDRESS:

ADDRESS: Sim & McBurney  
STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,671

FILING DATE: 07-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/337,483

FILING DATE: 08-NOV-1994

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/175,116

FILING DATE: 29-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/148,968

FILING DATE: 08-NOV-1993

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 911 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-474-671-95

Query Match 67.4%: Score 31: DB 3: Length 911:  
Best Local Similarity 75.0%: Pred. No. 2.8e+02:  
Matches 6: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

Oy 1 YOHDLRAY 8





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:50:17 ; Search time 35.29 Seconds  
(without alignments)  
21.783 Million cell updates/sec

Title: US-09-613-092a-6\_COPY\_3\_10

Perfect score: 46

Sequence: 1 YOHDLRAY 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	80.4	284	2 B83236	conserved hypotet
2	36	78.3	267	2 S52347	hypothetical prote
3	36	78.3	451	2 T15718	hypothetical prote
4	35	76.1	563	2 T04598	hypothetical prote
5	35	76.1	447	2 D71812	udp-n-acetylmuram
6	35	76.1	502	2 B75287	sensor histidine k
7	34	73.9	167	2 JTM065	fast skeletal trop
8	34	73.9	481	2 F96802	hypothetical prote
9	34	73.9	1030	2 T51452	hypothetical prote
10	34	73.9	1224	2 T25770	hypothetical prote
11	33	71.7	317	2 B37388	hypothetical prote
12	33	71.7	330	2 B97368	probable DNA-bindi
13	33	71.7	330	2 AB2586	hypothetical sugar
14	33	71.7	436	2 A52160	MSS51 protein - ye
15	33	71.7	488	2 T47787	hypothetical prote
16	33	71.7	490	2 A84861	probable amine oxi
17	33	71.7	516	2 D96682	protein FLE22.18 l
18	33	71.7	762	2 H83415	cis/trans isomeras
19	33	71.7	949	2 B32105	clathrin-associate
20	32	69.6	199	2 H90024	hypothetical prote
21	32	69.6	213	2 S13398	chloroamphenicol O-
22	32	69.6	261	2 S72750	triose-phosphate 1
23	32	69.6	354	2 AD0552	S-adenosylmethioni
24	32	69.6	356	2 B38530	S-adenosylmethioni
25	32	69.6	356	2 D85536	S-adenosylmethioni
26	32	69.6	356	2 H90685	S-adenosylmethioni
27	32	69.6	385	2 T20701	hypothetical prote
28	32	69.6	479	2 D70676	probable PE protei
29	32	69.6	727	2 T10616	hypothetical prote

30	32	69.6	812	2 S64929	hypothetical prote
31	32	69.6	1101	2 T20881	hypothetical prote
32	32	69.6	1139	1 PAVBRS	alpha-a protein -
33	32	69.6	2329	2 S44625	C50c3.6 protein -
34	31	67.4	98	2 G97041	hypothetical prote
35	31	67.4	104	2 B71180	hypothetical prote
36	31	67.4	133	2 D64609	conserved hypotet
37	31	67.4	158	2 G64885	ydar protein - Esc
38	31	67.4	158	2 B90817	probable phage rep
39	31	67.4	158	2 F85676	unknown protein en
40	31	67.4	260	2 E87577	glutamine cyclotra
41	31	67.4	263	1 A26147	egg-laying hormone
42	31	67.4	271	1 ONCAPA	egg-laying hormone
43	31	67.4	342	2 T14695	P2 protein homolog
44	31	67.4	348	2 G82645	alcohol dehydrogen
45	31	67.4	353	2 D83361	alcohol dehydrogen

## ALIGNMENTS

RESULT 1  
B83236  
conserved hypothetical protein PA3283 [imported] - Pseudomonas aeruginosa (strain PAO  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: B83236  
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Mickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
Nature 406, 959-964, 2000  
A>Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: B83236  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-284 <STO>  
A:Cross-references: GB:AE004750; GB:AE004091; NID:9949400; PIDN:AMG06671.1; GSPDB:GN  
A:Experimental source: strain PAO1  
C:Genetics:  
A:Gene: PA3283

Query Match 80.4%; Score 37; DB 2; Length 284;  
Best Local Similarity 85.7%; Pred. No. 7;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 QHDLRAY 8  
|||||  
Db 88 QHDLRAY 94

RESULT 2  
S52347  
hypothetical protein 1 - Lactobacillus leichmannii  
C:Species: Lactobacillus leichmannii  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999  
C/Accession: S52347  
R/Schenk-Groeninger, R.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S52347  
A:Accession: S52347  
A:Molecule type: DNA  
A:Residues: 1-267 <SCH>  
A:Cross-references: EMBL:X61869; NID:9666067; PIDN:CAA57458.1; PID:9666068  
A:Experimental source: DSM 20076

Query Match 78.3%; Score 36; DB 2; Length 267;  
Best Local Similarity 75.0%; Pred. No. 10;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 YOHDLRAY 8

Db 168 YNHDLR175 1|||||1

RESULT 3  
115718  
hypothetical protein C30G12.1 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15718  
R:Latreille, P.  
Submitted to the EMBL Data Library, July 1995  
A:Description: The sequence of *C. elegans* cosmid C30G12.  
A:Reference number: Z18393  
A:Accession: T15718  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-451 <LAT>  
A:Cross-references: EMBL:U21319; NID:q687832; PID:q687833; PIDN:AAC46670.1; CESP:C30G12.  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:C30G12.1  
A:Introns: 37/3; 71/2; 107/3; 153/2; 205/1; 289/1; 376/3

Query Match 78.3%; Score 36; DB 2; Length 451;  
Best Local Similarity 75.0%; Pred. No. 18;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8 1|||||1  
Db 62 YOHDLRAF 69

RESULT 4  
T04598  
hypothetical protein F23E13.180 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 15-Sep-2000  
C:Accession: T04598  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.  
Submitted to the Protein Sequence Database, March 1998  
A:Reference number: Z15378  
A:Accession: T04598  
A:Molecule type: DNA  
A:Residues: 1-563 <BEV>  
A:Cross-references: EMBL:AL022141  
A:Experimental source: cultivar Columbia; BAC clone F23E13  
C:Genetics:  
A:Map position: 4  
A:Introns: 80/1; 105/1; 116/3; 162/1; 192/1; 217/2; 243/3; 278/3; 327/3; 372/2; 467/3; 4  
A:Note: F23E13.180  
C:Superfamily: *Arabidopsis thaliana* hypothetical protein F23E13.180

Query Match 78.3%; Score 36; DB 2; Length 563;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8 1|||||1  
Db 273 YRHSIRAY 280

RESULT 5  
D71812  
udp-n-acetylglucosyl-tryptophan synthetase - *Helicobacter pylori* (strain J99)  
C:Species: *Helicobacter pylori*  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Apr-2000  
C:Accession: D71812  
R:All, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Voyts, G.F.;

Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p  
A:Reference number: A71800; MUID:99120557  
A:Accession: D71812  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-447 <ARN>  
A:Cross-references: GB:AE001561; GB:AE001439; NID:q415600; PIDN:AAD06968.1; PID:q415  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: murE  
C:Superfamily: UDP-N-acetylglucosylalanine-D-glutamate--2,6-diaminopimelate ligase

Query Match 76.1%; Score 35; DB 2; Length 447;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YOHDLR 6 1|||||1  
Db 238 YOHDLR 243

RESULT 6  
B75287  
sensor histidine kinase - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: B75287  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M. Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: B75287  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-502 <WH>  
A:Cross-references: GB:AE002064; GB:AE000513; NID:q6460134; PIDN:APF11875.1; PID:q646  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2328  
A:Map position: 1  
C:Superfamily: envZ protein; sensor histidine kinase homology

Query Match 76.1%; Score 35; DB 2; Length 502;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YOHDLR 6 1|||||1  
Db 310 YOHDLR 315

RESULT 7  
JW0065  
fast skeletal tropomyosin C - *Entosphenus japonicus*  
N:Alternate names: ftnC  
C:Species: *Entosphenus japonicus*  
C:Date: 13-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 29-Sep-1999  
C:Accession: JW0065  
R:Yusasa, H.J.; Cox, J.A.; Takagi, T.  
J. Biochem. 123, 1180-1190, 1998  
A:Title: Diversity of the tropomyosin C genes during chordate evolution.  
A:Reference number: JW0065; MUID:98265050  
A:Accession: JW0065  
A:Molecule type: mRNA  
A:Residues: 1-167 <YUA>  
A:Cross-references: DDBJ:AB008555; NID:q2589013; PIDN:BA23282.1; PID:q2589014  
C:Comment: This protein belongs to the EF-hand Ca<sup>2+</sup> binding protein family and functi  
C:Genetics:

A:introns: 110/2  
 C:superfamily: calmodulin; calmodulin repeat homology  
 C:Keywords: EF hand  
 F:20-52/Domain: calmodulin repeat homology <EF1>  
 F:56-88/Domain: calmodulin repeat homology <EF2>  
 F:96-128/Domain: calmodulin repeat homology <EF3>  
 F:132-164/Domain: calmodulin repeat homology <EF4>

Query Match 73.9%; Score 34; DB 2; Length 167;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YOHDLRAY 8  
 ||| |||  
 Db 10 QHDARAY 16

RESULT 8  
 F96802  
 hypothetical protein F2P24.7 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: F96802  
 R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maill, R.; Marziani, R.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; M01D:21016719  
 A:Accession: F96802  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-481 <STO>  
 A:Cross-references: GB:AE005173; NID:g11079489; PIDN:AA629201.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F2P24.7  
 A:Map position: 1

Query Match 73.9%; Score 34; DB 2; Length 481;  
 Best Local Similarity 62.5%; Pred. No. 48;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
 ||| |||  
 Db 94 YEHSVRAY 101

RESULT 9  
 T51452  
 hypothetical protein F2G14.190 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: T51452  
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew  
 submitted to the Protein Sequence Database, August 2000  
 A:Reference number: Z25394  
 A:Accession: T51452  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1030 <SAT>  
 A:Cross-references: EMBL:AL391146  
 A:Experimental source: cultivar Columbia; BAC clone F2G14  
 C:Genetics:  
 A:Map position: 5  
 A:introns: 29/3; 46/2; 74/2; 95/3; 147/1; 170/3; 205/3; 236/3; 261/3; 290/2; 359/2; 392/3

A>Note: F2G14\_190  
 C:superfamily: Saccharomyces cerevisiae hypothetical protein YLR410W

Query Match 73.9%; Score 34; DB 2; Length 1030;  
 Best Local Similarity 62.5%; Pred. No. 1,1e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
 ||| |||  
 Db 522 YRHDLKY 529

RESULT 10  
 T25770  
 hypothetical protein F46F11.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T25770  
 R:Pauley, A.; Gatlung, S.  
 submitted to the EMBL Data Library, February 1997  
 A:Description: The sequence of C. elegans cosmid F46F11.  
 A:Reference number: Z20083  
 A:Accession: T25770  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1224 <PAU>  
 A:Cross-references: EMBL:U88173; PIDN:ABA42264.1; GSPDB:GN00019; CESP:F46F11.1  
 A:Experimental source: strain Bristol N2; clone F46F11  
 C:Genetics:  
 A:Gene: CESP:F46F11.1  
 A:Map position: 1  
 A:introns: 16/3; 45/3; 84/3; 106/3; 189/3; 311/3; 429/3; 474/2; 523/1; 603/3; 742/3;

Query Match 73.9%; Score 34; DB 2; Length 1224;  
 Best Local Similarity 62.5%; Pred. No. 1,3e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
 ||| |||  
 Db 565 YRHDLKY 572

RESULT 11  
 B37388  
 probable DNA-binding protein 1B - Thermus aquaticus insertion sequence IS1000B  
 N:Alternate names: hypothetical protein 1B  
 C:Species: Thermus aquaticus  
 C>Date: 17-Apr-1993 #sequence\_revision 14-May-1993 #text\_change 08-Oct-1999  
 C:Accession: B37388; S27740  
 R:Ashby, M.K.; Bergquist, P.L.  
 Plasmid 24, 1-11, 1990  
 A>Title: Cloning and sequence of IS1000, a putative insertion sequence from Thermus  
 A:Reference number: A37388; M01D:91102148  
 A:Accession: B37388  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-317 <ASH>  
 A:Cross-references: EMBL:M33159; NID:g155103; PIDN:AAA27498.1; PID:g155109  
 A:Experimental source: strain HB8  
 A>Note: the source is designated as Thermus thermophilus  
 C:Genetics:  
 A:Mobile element: insertion sequence IS1000B  
 C:Keywords: DNA binding

Query Match 71.7%; Score 33; DB 2; Length 317;  
 Best Local Similarity 75.0%; Pred. No. 49;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
 ||| |||

Db 110 YHEDLRAY 117

RESULT 12

B97368

hypothetical sugar kinase slr0537 [imported] - Agrobacterium tumefaciens (strain C58, C6

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: B97368

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens

A:Reference number: A97359; PMID:11743194

A:Accession: B97368

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-330 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK65899.1; PID:g15154944; GSPDB:GN00169

C:Genetics:

A:Gene: AGR\_C\_118

A:Map position: circular chromosome

Query Match

Best Local Similarity 71.7%; Score 33; DB 2; Length 330;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YOHDLRA 7

Db 90 FOHDIRA 96

RESULT 13

AB2586

pfkB family carbohydrate kinase [imported] - Agrobacterium tumefaciens (strain C58, Dupc

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AB2586

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Bliddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AB2586

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-330 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAI41104.1; PID:g17738396; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu0079

A:Map position: circular chromosome

Query Match

Best Local Similarity 71.7%; Score 33; DB 2; Length 330;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YOHDLRA 7

Db 90 FOHDIRA 96

RESULT 14

S42160

MSS51 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein L8167.17; protein YLR203c

C:Species: Saccharomyces cerevisiae

C:Date: 03-May-1994 #sequence\_revision 07-Feb-1997 #text\_change 21-Jul-2000

C:Accession: S48554; S42160; S25344

R:Pauley, A.

submitted to the EMBL Data Library, September 1994

A:Description: The sequence of S. cerevisiae cosmid 8167.

A:Reference number: S48545

A:Accession: S48554

A:Molecule type: DNA

A:Residues: 1-436 <PAU>

A:Cross-references: EMBL:U14913; NID:G544497; PID:G544513; MIPS:YLR203c

R:Faye, G.; Simon, M.

Cell 32, 77-87, 1983

A:Title: Analysis of a yeast nuclear gene involved in the maturation of mitochondrial

A:Reference number: S42160; MUID:83128417

A:Accession: S42160

A:Molecule type: DNA

A:Residues: 1-209, 'M', 211-436 <FAY>

A:Cross-references: EMBL:J01487; NID:G172009; PIDN:AAA66926.1; PID:g172010

R:Simon, M.; della Seta, F.; Sor, F.; Faye, G.

Yeast 8, 559-567, 1992

A:Title: Analysis of the MSS51 region on chromosome XII of Saccharomyces cerevisiae.

A:Reference number: S25342; MUID:92397593

A:Accession: S25344

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-209, 'M', 211-436 <SIM>

A:Cross-references: GB:S43721; NID:G255246; PIDN:AA823218.1; PID:g255249

C:Genetics:

A:Gene: SGD:MSS51

A:Cross-references: SGD:S0004193; MIPS:YLR203c

A:Map position: 12R

C:Superfamily: Saccharomyces cerevisiae MSS51 protein

C:Keywords: mitochondrion

Query Match

Best Local Similarity 71.7%; Score 33; DB 2; Length 436;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YOHDLRA 7

Db 131 YEHDIRS 137

RESULT 15

T47787

hypothetical protein F17J16.100 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47787

R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzini, M.; Valle, G.; Mewes, H.W.; Rudd, S.

submitted to the Protein Sequence Database, April 2000

A:Accession: T47787

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-488 <DAN>

A:Cross-references: EMBL:AL163527

A:Experimental source: cultivar Columbia; BAC clone F17J16

C:Genetics:

A:Map position: 3

A:Introns: 14/1; 83/2; 121/2; 150/3; 179/2; 213/3; 246/2; 389/3

A:Note: F17J16.100

Query Match

Best Local Similarity 71.7%; Score 33; DB 2; Length 488;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8

Db 115 YDHDESY 122

Search completed: August 15, 2002, 11:50:18  
Job time: 244 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:58:16 ; Search time 18.61 Seconds

(without alignments)  
16.645 Million cell updates/sec

Title: US-09-613-092a-6\_COPY\_3\_10  
Perfect score: 46  
Sequence: 1 YOHDLRAY 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:.\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	78.3	451	YOB1_CAEEL	009355 caenorhabdi
2	35	76.1	447	MURE_HELPJ	0923c6 helicobacte
3	34	73.9	301	RL5_NEUCR	059953 neuropept
4	33	71.7	436	MS51_YEAST	P32335 saccharomyc
5	33	69.6	949	ALB1_RAT	P52303 ratu
6	32	69.6	213	CAT2_ECOLI	P26105 escherichia
7	32	69.6	261	TPIS_MYCLE	P46711 mycobacteri
8	32	69.6	356	QUBA_ECOLI	P21516 escherichia
9	32	69.6	385	KUPL_CAEEL	P34705 caenorhabdi
10	32	69.6	708	NICA_MOUSE	P57716 mus musculu
11	32	69.6	1139	VRNA_BSMV	P17595 barley stri
12	32	69.6	2329	YLJ6_CAEEL	P34369 caenorhabdi
13	31	67.4	158	RACR_ECOLI	P76062 escherichia
14	31	67.4	263	ELH1_APLPA	P17685 aplysia par
15	31	67.4	271	ELH1_APLCA	P01362 aplysia cal
16	31	67.4	432	FGI2_MOUSE	P12804 mus musculu
17	31	67.4	533	VIE2_AGRTE	P07544 agrobacteri
18	31	67.4	587	YNJ4_YEAST	P53933 saccharomyc
19	31	67.4	830	YKRT1_YEAST	P40850 saccharomyc
20	31	67.4	863	YP67_CAEEL	009216 caenorhabdi
21	31	67.4	911	TB11_NEIMB	009056 neisseria m
22	31	67.4	937	MSH2_NEUCR	013336 neuropept
23	30.5	66.3	749	VP4_ROTGA	004916 rotavirus (
24	30	65.2	146	YD34_MYCTU	010645 mycobacteri
25	30	65.2	259	MOTR_TREPA	007886 treponema p
26	30	65.2	342	YIIS_BACSU	P40332 bacillus su
27	30	65.2	423	YB12_SFV1	P29170 simian foam
28	30	65.2	403	YCD8_ECOLI	P31545 escherichia
29	30	65.2	467	AFCA_ARATH	P51566 arabidopsis
30	30	65.2	528	YNH7_YEAST	P53940 saccharomyc
31	30	65.2	552	YHNE_YEAST	P38799 saccharomyc
32	30	65.2	593	FA12_BOVIN	P58140 bos taurus
33	30	65.2	615	FA12_HUMAN	P00748 homo sapien

34	30	65.2	686	HMCT_HELPJ	Q92153 helicobacte
35	30	65.2	686	HMCT_HELPJ	059465 helicobacte
36	30	65.2	770	STA3_HUMAN	P40763 homo sapien
37	30	65.2	770	STA3_MOUSE	P42227 mus musculu
38	30	65.2	770	STA3_RAT	P52631 ratu
39	30	65.2	869	AMPN_ECOLI	P04825 escherichia
40	30	65.2	942	HEX_ADEG1	P42671 avian adeno
41	30	65.2	1341	YL78_YEAST	005854 saccharomyc
42	30	65.2	1750	Y832_METJA	058242 methanococc
43	30	65.2	2388	SPCP_RAT	099498 ratu
44	30	65.2	2390	SPCP_HUMAN	015020 homo sapien
45	29	63.0	91	XPR7_BACLI	Q99166 bacillus 11

## ALIGNMENTS

```
RESULT 1
YOB1_CAEEL          STANDARD;          PRT;          451 AA.
AC 009255;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 51.0 kDa protein C30G12.1 in chromosome II.
GN C30G12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC SRRAIN-BRISTOL NZ;
RA Latreille P.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: U21319; AAC46670.1; -.
DR WormRep: C30G12.1; CE01832.
KW Hypothetical protein.
SQ
SEQUENCE 451 AA; 51017 MW; 6AAAFD32675ICCI CRC64;
```

Query Match 78.3%; Score 36; DB 1; Length 451;  
Best Local Similarity 75.0%; Pred. No. 4.5;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YOHDLRAY 8  
||||| :  
Db 62 YOHDLRAF 69

RESULT 2  
MURE\_HELPJ  
ID MURE\_HELPJ STANDARD; PRT; 447 AA.  
AC 0923c6;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE UDP-N-acetylmuramoylalanine-D-glutamate--2,6-diaminopimelate ligase  
DE (EC 6.3.2.13) (UDP-N-acetylmuramyl-tripeptide synthetase) (Meso-  
DE diaminopimelate-adding enzyme) (UDP-MurNAc-tripeptide synthetase).  
GN MURE OR JHP1387.  
OS Helicobacter pylori J99 (Campylobacter pylori J99).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.

```

ON NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A. PubMed-9923682;
RX MEDLINE=99120557;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D., Noonan B., Guild B.C., deJonge B.L., Carnel G.,
RA Tummino P.J., Caruso A., Uria-Nikelsen M., Mills D.M., Ives C.,
RA Gibson S., Merberg D., Mills S.D., Jang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- FUNCTION: CELL WALL FORMATION. DIAMINOPIMELIC ACID ADDING ENZYME
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamate + meso-2,6-diaminoheptanate = ADP + phosphate +
CC UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-
CC diaminoheptanate.
CC -1- PATHWAY: PEPTIDOLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE MURCEDE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE001561; AAD06968.1; -
DR InterPro: IPR000713; Mur_Llgase.
DR InterPro: IPR004101; Mur_Llgase.C.
DR Pfam: PF02875; Mur_Llgase_1.
DR Pfam: PF02875; Mur_Llgase_C_1.
KW Peptidoglycan synthesis; Cell wall; Cell division; Llgase;
KW Atp-binding; Complete proteome.
FT NP_BIND 74 80 ATP (POTENTIAL).
SQ SEQUENCE 447 AA; 50660 MW; 639AB9D18F3740BD CRC64;

Query Match 76.1%; Score 35; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YOHDLR 6
DB 238 YOHDLR 243

RESULT 3
RL5_NEUCR
ID RL5_NEUCR STANDARD; PRT; 301 AA.
AC O59953;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60S ribosomal protein L5 (CPR4).
GN CRP-4.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OC NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RC de la Serna I.L., Cujeç T.P., Shi Y., Tyler B.M.;
RT "Non-coordinate regulation of 5S rRNA genes and the gene encoding the
RT 5S rRNA-binding ribosomal protein homolog in Neurospora crassa.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN BINDS 5S RNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
CC -----

```

Query	Match	Best Local	Similarity	Score	DB	Length	301	Matches	5	Conservative	2	Mismatches	1	Indels	0	Gaps	0
QY	1	YOHDLRAY	8														
Db	79	YSHEKAK	86														
RESULT	4																
MS51_YEAST																	
ID	MS51_YEAST	STANDARD;	PRT;	436	AA.												
AC	P32335.																
DT	01-OCT-1993	(Rel. 27, Created)															
DT	01-OCT-1993	(Rel. 27, Last sequence update)															
DT	01-OCT-1996	(Rel. 34, Last annotation update)															
DE	MS551 protein.																
GN	MS551 OR YLR1203C OR L8167.17.																
OS	Saccharomyces cerevisiae (Baker's yeast).																
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;																
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.																
OX	NCBI_TaxID=4932;																
LN	[1]																
RP	SEQUENCE FROM N.A.																
RP	STRAIN=AB320;																
RX	MEDLINE=83129417; PubMed=6297789;																
RA	Faye G., Simon M.;																
RT	"Analysis of a yeast nuclear gene involved in the maturation of																
RT	mitochondrial pre-messenger RNA of the cytochrome oxidase subunit																
RT	I.";																
RL	Cell 32:77-87(1983).																
RL	[2]																
RP	SEQUENCE FROM N.A.																
RP	STRAIN=DD273-10B/A;																
RX	MEDLINE=92397593; PubMed=1523888;																
RA	Simon M., della Seta F., Sor F., Faye G.;																
RT	"Analysis of the MS551 region on chromosome XII of Saccharomyces																
RT	cerevisiae.";																
RL	Yeast 8:559-567(1992).																
RL	[3]																
RP	SEQUENCE FROM N.A.																
RP	STRAIN=S288C / AB972;																
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,																
RA	Favellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,																
RA	Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,																
RA	Johnston L., Langston Y., Latreille P., Mardis E., Meneses S.,																
RA	Miller N., Nian M., Pauley A., Peluso D., Rifken L., Riles L.,																
RA	Tatch S., Travakis E., Vaidin M., Vignati D., Wilcox L., Wilson R.,																
RA	Wohlman P., Wakerston R.;				</												



RL Mol. Genet. 224:111-118(1990).  
 CC -1- FUNCTION: NECESSARY FOR THE CORRECT MATURATION OF THE  
 CC MITOCHONDRIAL PRE-MRNA FOR CYTOCHROME OXIDASE SUBUNIT 1 (COX1). IT  
 CC IS PROBABLY AN INITIATION FACTOR SPECIFIC FOR THE TRANSLATION OF  
 CC THE COX1 MRNA.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J01487; AAA66926.1; -  
 DR EMBL: S43721; AAB23218.1; -  
 DR EMBL: U14913; AAB67438.1; -  
 DR PIR: S25344; S25344.  
 DR PIR: S42160; S42160.  
 DR PIR: S48554; S48554.  
 DR SGD: S0004193; MSS51.  
 KW mRNA processing; mRNA splicing; Mitochondrion.  
 FT VARIANT 210 210 M -> I (IN STRAINS D273-10B/A AND AB972).  
 FT VARIANT 339 339 G -> S (IN STRAIN D273-10B/A).  
 FT VARIANT 375 375 T -> M (IN STRAIN D273-10B/A).  
 FT VARIANT 397 397 R -> K (IN STRAIN D273-10B/A).  
 SQ SEQUENCE 436 AA; 50898 MW; B7616DDBAFC32FBI CRC64;

Query Match 71.7%; Score 33; DB 1; Length 436;  
 Best Local Similarity 71.4%; Pred. No. 18;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 YQHDLR 7  
 DB 131 YQHDLR 137

RESULT 5  
 AIBL\_RAT STANDARD: PRT; 949 AA.  
 ID AIBL\_RAT  
 AC P52303;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Adapter-related protein complex 1 beta 1 subunit (beta-adaptin 1)  
 DE (Plasma membrane adaptor HA2/AP2 adaptin beta subunit) (Clathrin  
 assembly protein complex 2 beta large chain) (AP105A).  
 GN APTB1 OR ADTB1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=89202379; PubMed=2495531;  
 RA Kirchhausen T., Nathanson K.L., Matsui W., Vaisberg A., Chow E.P.,  
 RA Burne C., Keen J.H., Davis A.E.;  
 RT "Structural and functional division into two domains of the large  
 RT (100- to 115-kDa) chains of the clathrin-associated protein complex  
 RT AP-2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2612-2616(1989).  
 CC -1- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES  
 CC WHICH LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-  
 CC ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE  
 CC CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION  
 CC AND CONCENTRATION. BETA ADAPTIN IS A SUBUNIT OF THE PLASMA  
 CC MEMBRANE ADAPTOR.  
 CC -1- SUBUNIT: ASSEMBLY PROTEIN COMPLEX 2 (AP-2) IS A HETEROTETRAMER  
 CC COMPOSED OF TWO LARGE CHAINS (ALPHA AND BETA), A MEDIUM CHAIN  
 CC (AP50) AND A SMALL CHAIN (AP17).

CC -1- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE  
 CC CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNITS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M77245; AAA40807.1; -  
 DR InterPro: IPR002553; Adaptin\_N.  
 DR InterPro: IPR001121; G\_adapt\_C.  
 DR Pfam: PF01602; Adaptin\_N.1.  
 DR Pfam: PF02883; Alpha\_adaptinC2.1.  
 KW Coated pits.  
 FT DOMAIN 576 728 PRO-RICH (STALK REGION).  
 SQ SEQUENCE 949 AA; 104568 MW; CC5AE34EBAACD44 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 949;  
 Best Local Similarity 71.4%; Pred. No. 42;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 QHDLRAY 8  
 DB 937 QHDLRAY 943

RESULT 6  
 CAT2\_ECOLI STANDARD: PRT; 213 AA.  
 ID CAT2\_ECOLI  
 AC P22615;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Chloramphenicol acetyltransferase II (EC 2.3.1.28) (CAT-II).  
 GN CMLA.  
 OS Escherichia coli.  
 OG Plasmid Incw psa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=K12;  
 RX MEDLINE=91097467; PubMed=2268278;  
 RA Murray I.A., Martinez-Suarez J.V., Close T.J., Shaw W.V.;  
 RT "Nucleotide sequences of genes encoding the type II chloramphenicol  
 RT acetyltransferases of Escherichia coli and Haemophilus influenzae,  
 RT which are sensitive to inhibition by thiol-reactive reagents.";  
 RL Biochem. J. 272:505-510(1990).  
 CC -1- FUNCTION: THIS ENZYME IS AN EFFECTOR OF CHLORAMPHENICOL RESISTANCE  
 CC IN BACTERIA.  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + chloramphenicol = CoA +  
 CC chloramphenicol 3-acetate.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- MISCELLANEOUS: TYPE II CHLORAMPHENICOL ACETYLTANSFERASES ARE  
 CC SENSITIVE TO INHIBITION BY THIOL-REACTIVE REAGENTS. THE  
 CC INACTIVATION OCCURS AS A RESULT OF CHEMICAL MODIFICATION OF  
 CC CYS-26.  
 CC -1- SIMILARITY: BELONGS TO THE CHLORAMPHENICOL ACETYLTANSFERASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: X53796; CAA37805.1; -  
 DR PIR: S13398; S13398.  
 DR HSSP: P00484; 10CA.  
 DR InterPro: IPR001707; CAT.  
 DR Pfam: PF00302; CAT: 1.  
 DR ProDom: PD002660; CAT: 1.  
 DR PROSITE: PS00100; CAT: 1.  
 KM Antibiotic resistance: Transferase: Acyltransferase; Plasmid.  
 FT ACT\_SITE 189 189 BY SIMILARITY  
 SQ SEQUENCE 213 AA; 24778 MW; 2E0D7F150A730AE7 CRC64;  
 -----  
 Query Match 69.6%; Score 32; DB 1; Length 213;  
 Best Local Similarity 62.5%; Pred. No. 13;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 -----  
 QY 1 YOHDLRAY 8  
 Db 122 YOHDLRLF 129  
 -----  
 RESULT 7  
 TPIS\_MYCLE STANDARD; PRT; 261 AA.  
 AC P46711;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Triosephosphate isomerase (EC 5.3.1.1) (TIM).  
 GN TPJA OR TPI OR ML0572 OR B1496\_C1\_127.  
 OS Mycobacterium lepreae.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R., Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 -----  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TN;  
 RA MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,  
 RA Holtroyd S., Hornsby T., Jagels K., Jancic C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor S., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate -> glyceralone  
 phosphate.  
 CC -1- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TRIOSPHOSPHATE ISOMERASE FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: U00013; AAA17115.1; -  
 DR EMBL: AL583919; CAC30080.1; -  
 DR HSSP: P00943; 2B7M.

DR Leproma; M0572; -  
 DR InterPro: IPR000652; Trioseph\_isomerase.  
 DR Pfam: PF00121; TIM; 1.  
 DR ProDom: PD001005; Trioseph\_isomerase; 1.  
 DR PROSITE: PS00171; TIM; 1  
 KM Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;  
 KW Pentose shunt; Complete proteome.  
 FT ACT\_SITE 100 100 BY SIMILARITY.  
 FT ACT\_SITE 172 172 BY SIMILARITY.  
 SQ SEQUENCE 261 AA; 27557 MW; 598B5CDA6BBDB822D CRC64;  
 -----  
 Query Match 69.6%; Score 32; DB 1; Length 261;  
 Best Local Similarity 85.7%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 -----  
 QY 2 QHDLRAY 8  
 Db 73 QHDLGAY 79  
 -----  
 RESULT 8  
 QUEA\_ECOLI STANDARD; PRT; 356 AA.  
 AC P21516;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE S-adenosylmethionine:tRNA ribosyltransferase-isomerase (EC 5.4.1.1) (Q)  
 DE (Quenosine biosynthesis protein queA).  
 GN QUESA OR B0405 OR Z0504 OR ECS0456.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA MEDLINE=91177815; PubMed=1706703;  
 RA Reuter K., Slany R., Ullrich F., Kersten H.;  
 RT "Structure and organization of Escherichia coli genes involved in  
 RT biosynthesis of the deazaguanine derivative queuine, a nutrient  
 RT factor for eukaryotes."  
 RL J. Bacteriol. 173:2256-2264(1991).  
 -----  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 -----  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,  
 RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,  
 RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 -----  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 RA MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Godtcek E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blatner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
 RL Nature 409:529-533(2001).

```

RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-0157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
  Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
  Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
  O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE-93349860; PubMed-8347586;
RA Slany R.K., Boesl M., Crain P.F., Kersten H.;
RT "A new function of S-adenosylmethionine: the ribosyl moiety of AdoMet
  is the precursor of the cyclopentenol moiety of the tRNA wobble
  base sequence."
RL Biochemistry 32:7811-7817(1993).
CC -1- FUNCTION: SYNTHESIZES OO FROM PREQ1 IN A SINGLE S-
  ADEOSYLMETHIONINE- (ADOMET-) REQUIRING STEP. THE RIBOSYL MOIETY
  OF ADOMET IS TRANSFERRED AND ISOMERIZED TO THE EPOXYCYClopENTANE
  RESIDUE OF OO.
CC -1- PATHWAY: INVOLVED IN BIOSYNTHESIS OF THE DEAZAGUANINE DERIVATIVE
  QUEUOSINE, WHICH IS USUALLY PRESENT IN TRNAS SPECIFIC FOR ASN,
  ASP, HIS, AND TYR.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE QUEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37702; AAA16114.1; -
DR EMBL: AE000147; AAC73508.1; -
DR EMBL: U82664; AAB40161.1; -
DR EMBL: AE005220; AAG54752.1; -
DR EMBL: AF002551; BAB3879.1; -
DR PIR: B38530; B38530.
DR PIR: S35451; S35451.
DR EcoGene: EG10812; queA.
DR InterPro: IPR003699; Queuosine_synth.
DR Pfam: PF02547; Queuosine_synth.1.
KW Queuosine biosynthesis; Transferase; Isomerase; Complete proteome.
SQ SEQUENCE 356 AA; 39430 MW; 47F7F0090812DCFC3 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 356;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8
DB 317 YOHDMAY 324

RESULT 9
KUP1_CAEEL STANDARD; PRT; 385 AA.
AC P34705;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein kup-1.
GN KUP-1 OR F10C2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
  Rhabditidae; Pelodidae; Caenorhabditis.

```

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-93258816; PubMed-8098272;
RA Spleth J., Brooke G., Kuersten S., Lea K., Blumenthal T.;
RT "Operons in C. elegans: polycistronic mRNA precursors are processed
  by trans-splicing of SL2 to downstream coding regions."
RL Cell 73:521-532(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA White S.;
RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L12247; AAA28102.1; -
DR EMBL: Z81497; CAB04080.1; -
DR WormPeP: F10C2.2; CE09306.
SQ SEQUENCE 385 AA; 44188 MW; A023126D4E10A23B CRC64;

```

```

Query Match 69.6%; Score 32; DB 1; Length 385;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8
DB 281 YEHDTREY 288

RESULT 10
NICAMOUSE STANDARD; PRT; 708 AA.
AC P57716;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nicastatin precursor.
GN NCSTN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20445163; PubMed-10993067;
RA Yu G., Nishimura M., Arawaka S., Levitan D., Zhang L., Tandon A.,
  Song Y.-O., Rogava E., Chen F., Kawarai T., Supala A., Levesque L.,
  Yu H., Yang D.S., Holmes E., Milman P., Zhang Y., Zhang D.M., Xu D.H.,
  RA Sato C., Rogava E., Smith M., Janus C., Zhang Y., Aebersold R.,
  RA Farrer L.S., Sorbi S., Brunl A., Fraser P.E., St George-Hyslop P.H.;
RT "Nicastatin modulates presenilin-mediated notch/glp-1 signal
  transduction and betaAPP processing."
RL Nature 407:48-54(2000).
CC -1- FUNCTION: MAY FUNCTION IN THE CYTOPLASMIC PARTITIONING AND
  INTRAMEMBRANEOUS PROTEOLYSIS OF PROTEINS. TOGETHER WITH PRESENILIN
  IS INVOLVED IN GAMMA-SECRETASE MEDIATED PROTEOLYTIC PROCESSING OF
  AMYLOID PRECURSOR PROTEIN (APP). MAY PLAY A ROLE IN NOTCH
  SIGNALING (BY SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH PSEN1 AND PSEN2 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE NICASTATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF240469; AAG11413.1; -  
 DR MGD: MGI:1891700; Ncstn.  
 KW Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 708 NICASTRIN.  
 FT DOMAIN 28 668 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 669 689 POTENTIAL.  
 FT DOMAIN 690 708 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 572 572 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 579 579 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 708 AA; 78489 MW; EB7B5FD9C9D3DECI CRC64;

Query Match 69.6%; Score 32; DB 1; Length 708;  
 Best Local Similarity 71.4%; Pred. No. 49;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 QHDLRAY 8  
 DB 538 KHDLSY 544

RESULT 11  
 VRNA\_BSMV STANDARD; PRT; 1139 AA.  
 AC P17595;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Alpha-A protein.  
 OS Barley stripe mosaic virus (BSMV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Hordeiviruses.  
 NC NCB1\_TaxID=12327;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC PV42;  
 RX MEDLINE=89268457; Pubmed=2728343;  
 RA Gustafson G., Armour S.L., Gamboa G.C., Burgett S.G., Shepherd J.W.;  
 RT "Nucleotide sequence of barley stripe mosaic virus RNA alpha: RNA  
 RT alpha encodes a single polypeptide with homology to corresponding  
 RT proteins from other viruses."  
 RL Virology 170:370-377(1989).  
 CC -I- SIMILARITY: TO THE TMV 126 KDA PROTEIN, AND WITH RELATED PEPTIDES  
 CC FROM OTHER VIRUSES. ITS C-TERMINUS HAS LIMITED HOMOLOGY WITH THE  
 CC 58 KDA (BETA-B) PROTEIN FROM BSMV.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: J04342; AAA46336.1; -  
 DR PIR: J40109; PAVBBS.  
 DR InterPro: IPR002588; V\_methyltransf.  
 DR InterPro: IPR000606; Viral\_helicase1.  
 DR Pfam: PF01443; Viral\_helicase1.  
 DR Pfam: PF01660; Vmethyltransf; 1.  
 KW ATP-binding.  
 FT NP\_BIND 838 845 ATP (POTENTIAL).  
 FT SEQUENCE 1139 AA; 129627 MW; 24B1500FD05362D9 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1139;  
 Best Local Similarity 62.5%; Pred. No. 81;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YQHDRLAY 8  
 DB 294 YSHDLKDY 301

RESULT 12  
 YLJ6\_CAEEL STANDARD; PRT; 2329 AA.  
 AC P34369;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 272.0 kDa protein C50C3.6 in chromosome III.  
 CN C50C3.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NC NCB1\_TaxID=62339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; Pubmed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Lacroille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Winstock L., Wilkinson-Sprat J.,  
 RA Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 CC -I- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -I- SIMILARITY: STRONG. TO YEAST PRP8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L14433; AAA2797.1; -  
 DR PIR: S44625; S44625.  
 DR WormPep: C50C3.6; CE00122.  
 DR InterPro: IPR000555; MOV34.  
 DR Pfam: PF01398; MOV34; 1.  
 DR SMART: SM00232; JAB\_MPN; 1.  
 KW Hypothetical protein; Nuclear protein; RNA-binding.  
 SQ SEQUENCE 2329 AA; 272025 MW; 626F9F97C800960C CRC64;

Query Match 69.6%; Score 32; DB 1; Length 2329;  
 Best Local Similarity 62.5%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YQHDRLAY 8  
 1:111111  
 DB 71 YRHDRKRVY 78

## RESULT 13

RACR\_ECOLI STANDARD; PRT; 158 AA.  
 AC P76062;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Rac prophage repressor.  
 GN RACR OR B1356.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP FUNCTION.  
 RA Clark A.J.;  
 RL Unpublished observations (JUN-1998).  
 CC -1- FUNCTION: REPRESSOR PROTEIN FOR RAC PROPHAGE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE000233; AAC74438.1; -.  
 DR Ecogene; E613360; RACR.  
 KW Transcription regulation; Repressor; DNA-binding; Complete proteome.  
 SQ SEQUENCE 158 AA; 17663 MW; 110E900ED3847DDF CRC64;

Query Match 67.4%; Score 31; DB 1; Length 158;  
 Best Local Similarity 83.3%; Pred. No. 16;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 HDRLAY 8  
 1:111111  
 DB 124 HDVRLAY 129

## RESULT 14

ELH1\_APLPA STANDARD; PRT; 263 AA.  
 AC P17685;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE ELH type 1 precursor [Contains: Alpha-bag cell peptide (Alpha-BCP);  
 DE Beta-bag cell peptide (Beta-BCP); Gamma-bag cell peptide (Gamma-BCP);  
 DE Egg-laying hormone (ELH); Acidic peptide].  
 GN ELH1.  
 OS Aplysia parvula (Little sea hare).

OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;  
 OC Aplysidae; Aplysia.  
 OX NCBI\_TaxID=6503;

RP SEQUENCE FROM N.A.  
 RX MEDLINE=86280675; PubMed=3734873;  
 RA Nambu J.R., Scheller R.H.;  
 RT "Egg-laying hormone genes of Aplysia: evolution of the ELH gene  
 RT family";  
 RL J. Neurosci. 6:2026-2036(1986)

-1- FUNCTION: ELH ACTS AS A NEUROTRANSMITTER LOCALLY, UPON NEURONS OF  
 CC THE ABDOMINAL GANGLION AND AS A HORMONE BY DIFFUSING INTO THE  
 CC CIRCULATING HEMOLYMPH AND MODULATING THE ACTIVITY OF OTHER ORGANS.  
 CC IT SPECIFICALLY CAUSES CONTRACTION OF SMOOTH MUSCLE IN THE  
 CC OVTESTIS AND EXPULSION OF THE EGG STRING.  
 CC -1- FUNCTION: ALPHA-BCP DECREASES THE ACTIVITY OF A CLUSTER OF NEURONS  
 CC IN THE LEFT UPPER QUADRANT OF THE ABDOMINAL GANGLION.  
 CC -1- FUNCTION: BETA-BCP SPECIFICALLY EXCITES 2 NEURONS, L1 AND R1, IN  
 CC THE ABDOMINAL GANGLION.  
 CC -1- TISSUE SPECIFICITY: BAG CELL NEURONS.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: M30048; AAA27765.1; -.  
 DR PIR: A26147; A26147.  
 DR InterPro: IPR003424; ELH.  
 DR Pfam: PF02323; ELH; 1.  
 KW Hormone; Cleavage on pair of basic residues; Amidation; Neuropeptide;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 29  
 FT PEPTIDE 88 92 BY SIMILARITY.  
 FT PEPTIDE 95 99 BETA-BAG CELL PEPTIDE.  
 FT PEPTIDE 142 150 GAMMA-BAG CELL PEPTIDE.  
 FT PEPTIDE 198 233 ALPHA-BAG CELL PEPTIDE.  
 FT PEPTIDE 237 263 EGG-LAYING HORMONE.  
 FT MOD.RES 233 253 ACIDIC PEPTIDE.  
 FT MOD.RES 233 253 AMIDATION (G-234 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 263 AA; 29676 MW; E0C69FE3E5FE29 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 263;  
 Best Local Similarity 85.7%; Pred. No. 27;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YQHDRLA 7  
 1:111111  
 DB 136 YQDRLA 142

## RESULT 15

ELH1\_APLCA STANDARD; PRT; 271 AA.  
 AC P01362;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE ELH precursor [Contains: Alpha-bag cell peptide (Alpha-BCP); Beta-bag  
 DE cell peptide (Beta-BCP); Gamma-bag cell peptide (Gamma-BCP); Egg-  
 DE laying hormone (ELH); Acidic peptide].  
 OS Aplysia californica (California sea hare).  
 OC Eukaryota; Metazoa; Mollusca; gastropoda; Opisthobranchia; Anaspidae;  
 OC Aplysidae; Aplysia.  
 OX NCBI\_TaxID=6500;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83129416; PubMed=6687446;  
 RA Scheller R.H., Jackson J.F., McAllister L.B., Rothman B.S.,  
 RA Mayeri E., Axel R.;

RT "A single gene encodes multiple neuropeptides mediating a stereotyped  
 RT behavior.";  
 RL Cell 32:7-22(1983).  
 RN [2]  
 RP SEQUENCE OF 150-158.  
 RA Rothman B.S., Mayeri E., Brown R.O., Yuan P.-M., Shively J.E.;  
 RT "Primary structure and neuronal effects of alpha-bag cell peptide, a  
 RT second candidate neurotransmitter encoded by a single gene in bag cell  
 RT neurons of Aplysia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:5753-5757(1983).  
 RN [3]  
 RP SEQUENCE OF 206-241.  
 RX MEDLINE-80101672; PubMed-293751;  
 RA Chiu A.Y., Hunkapiller M.W., Heller E., Stuart D.K., Hood L.E.,  
 RA Strumwasser F.;  
 RT "Purification and primary structure of the neuropeptide egg-laying  
 RT hormone of Aplysia californica.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 76:6656-6660(1979).  
 RN [4]  
 RP PROCESSING.  
 RX MEDLINE-87168672; PubMed-3549995;  
 RA Newcomb R., Scheller R.H.;  
 RT "Proteolytic processing of the Aplysia egg-laying hormone and R3-14  
 RT neuropeptide precursors.";  
 RL J. Neurosci. 7:854-863(1987).  
 CC -1- FUNCTION: ELH ACTS AS A NEUROTRANSMITTER LOCALLY, UPON NEURONS OF  
 CC THE ABDOMINAL GANGLION AND AS A HORMONE BY DIFFUSING INTO THE  
 CC CIRCULATING HEMOLYMPH AND MODULATING THE ACTIVITY OF OTHER ORGANS.  
 CC IT SPECIFICALLY CAUSES CONTRACTION OF SMOOTH MUSCLE IN THE  
 CC OVOTESTIS AND EXPULSION OF THE EGG STRING.  
 CC -1- FUNCTION: ALPHA-BCP DECREASES THE ACTIVITY OF A CLUSTER OF NEURONS  
 CC IN THE LEFT UPPER QUADRANT OF THE ABDOMINAL GANGLION.  
 CC -1- FUNCTION: BETA-BCP SPECIFICALLY EXCITES 2 NEURONS, L1 AND R1, IN  
 CC THE ABDOMINAL GANGLION.  
 CC -1- TISSUE SPECIFICITY: BAG CELL NEURONS.  
 CC -1- SIMILARITY: TO ATRIAL GLAND PEPTIDE A AND B PRECURSORS.  
 DR PIR: A01632; ONCAPA.  
 DR InterPro: IPR003424; ELH.  
 DR Pfam: PF02323; ELH; 1.  
 DR Hormone; Cleavage on pair of basic residues; Amidation; Neuropeptide;  
 KW Signal.  
 FT SIGNAL. 1 34  
 FT PROPEP 35 93 POTENTIAL.  
 FT PEPTIDE 96 100 BETA-BAG CELL PEPTIDE.  
 FT PEPTIDE 103 107 GAMMA-BAG CELL PEPTIDE.  
 FT PROPEP 110 148  
 FT PEPTIDE 150 158 ALPHA-BAG CELL PEPTIDE.  
 FT PROPEP 162 203  
 FT PEPTIDE 206 241 EGG-LAYING HORMONE.  
 FT PEPTIDE 245 271 ACIDIC PEPTIDE.  
 FT MOD\_RES 241 241 AMIDATION (G-242 PROVIDE AMIDE GROUP).  
 FT SIMILAR 206 241 TO LARGE SUBUNIT OF CALIFIN A, B AND  
 FT C OF APLYSIA CALIFORNICA.  
 FT TO SMALL SUBUNIT OF CALIFIN A, B AND  
 FT C OF APLYSIA CALIFORNICA.  
 FT SIMILAR 245 271  
 FT SEQUENCE 271 AA; 30827 MM; FB2F1EBE3B9113A6 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 271;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 YQHDLRA 7  
 11111  
 Db 144 YQDRLRA 150

Search completed: August 15, 2002, 11:58:17  
 Job time: 517 sec



RESULT 2  
Q9TU16 PRELIMINARY; PRT: 1763 AA.  
Q9TU16  
AC Q9TU16  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Apolipoprotein B (FRAGMENT).  
OS Monodelphis domestica (Short-tailed grey opossum).  
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.  
OX NCBI\_TaxID=13616;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fujino T., Navaratnam N., Jarmuz A., von Haeselaer A., Scott J.;  
RT "C to U editing of apolipoprotein B mRNA in marsupials :  
RT Identification and characterisation of APOBEC-1 from the American  
RT opossum Monodelphis domestica.";  
RL Nucleic Acids Res. 0:0-0(1999).  
DR EMBL; AB027413; BAA86052.1; -.  
DR InterPro; IPR001747; Vitellogenin\_N.  
DR Pfam; PF01347; Vitellogenin\_N.1.  
KW Lipoprotein.  
FT NON\_TER  
SO SEQUENCE 1763 AA; 192330 MW; DBE8F87ED2D6289 CRC64;

Query Match 80.4%; Score 37; DB 6; Length 1763;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
1 111111  
DB 1576 YDHDLQAY 1583

RESULT 3  
Q48706 PRELIMINARY; PRT: 267 AA.  
Q48706  
AC Q48706  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ORF1 PROTEIN.  
GN ORF1.  
OS Lactobacillus leichmannii.  
OC Bacteria; Firmicutes; Bacillus/clostridium group; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=28039;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-DSM 20076;  
RA Schenk-Groeninger R.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X81869; CAA57458.1; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN.1.  
SQ SEQUENCE 267 AA; 29995 MW; 29D606B935C559AB CRC64;

Query Match 78.3%; Score 36; DB 2; Length 267;  
Best Local Similarity 75.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
1 111111  
DB 168 YNHDLRTY 175

RESULT 4  
023245 PRELIMINARY; PRT: 408 AA.  
ID 023245

AC 023245;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 45.3 KDA PROTEIN (FRAGMENT).  
GN CTA10.1020.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Terryn N., Vos P., Heijnen L., Mewes H.W., Mayer K.F.X.,  
RA Schueller C.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z99708; CAB16854.1; -.  
KW Hypothetical protein.  
FT NON\_TER  
SO SEQUENCE 408 AA; 45344 MW; 64DDCCDD3889AACC CRC64;

Query Match 78.3%; Score 36; DB 10; Length 408;  
Best Local Similarity 75.0%; Pred. No. 37;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
1 1 11111  
DB 365 YRHSRLAY 372

RESULT 5  
O65520 PRELIMINARY; PRT: 563 AA.  
O65520  
AC O65520  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 62.5 KDA PROTEIN.  
GN F23E13.180 OR AT4G36290.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Hohnel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.,  
RA Schueller C.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL022141; CAA18135.1; -.  
DR EMBL; AL161589; CAB80300.1; -.  
KW Hypothetical protein.  
SO SEQUENCE 563 AA; 62526 MW; BE176FF1601C4C8D CRC64;

Query Match 78.3%; Score 36; DB 10; Length 563;  
Best Local Similarity 75.0%; Pred. No. 52;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



OY 1 YOHDLRAY 8  
1:11111  
DB 273 YRHSRLRAY 280

RESULT 6  
O9C5D9 PRELIMINARY; PRT: 635 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE HYPOTHEICAL 70.8 KDA PROTEIN.  
GN ATAG36290.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
RA Carninci P., Chen H., Chauk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Full length cDNA of gene AT4g36290 (GI:7270582)."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF360314; AAK26024.1; -  
DR InterPro: IPR003594; HATPase\_C.  
DR SMART; SM00387; HATPase\_C; 1.  
KW Hypothetical protein.  
FT VARIANT 77  
SQ SEQUENCE 635 AA; 70819 MW; 30D2718E8CBAF53 CRC64;

Query Match 78.3%; Score 36; DB 10; Length 635;  
Best Local Similarity 75.0%; Pred. No. 60;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
1:11111  
DB 330 YRHSRLRAY 337

RESULT 7  
O93XR9 PRELIMINARY; PRT: 756 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE BG55 PROTEIN.  
GN BG55.  
OS Bruguiera gymnorhiza.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Malpighiales; Rhizophoraceae; Bruguiera.  
OX NCBI\_TaxID=39984;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAF;  
RA Banzai T., Hershkovits G., Katcoff D.J., Hanagata N., Dubinsky Z.,  
RA Karube I.;  
RT Identification of mRNA transcripts differentially expressed in  
RT response to high salinity by means of differential display in the  
RT mangrove plant, Bruguiera gymnorhiza.  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB061795; BAB55653.1; -  
SQ SEQUENCE 756 AA; 84482 MW; 1597F3D46C3E0809 CRC64;

Query Match 78.3%; Score 36; DB 10; Length 756;  
Best Local Similarity 85.7%; Pred. No. 72;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 YOHDLRA 7  
111111  
DB 480 YOHDLRA 486

RESULT 8  
O987E1 PRELIMINARY; PRT: 345 AA.

DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DE MLT7088 PROTEIN.  
GN MLT7088.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti."  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP003011; BAB53262.1; -  
DR InterPro: IPR001173; Glycos\_transf\_2.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 345 AA; 39450 MW; 131B52D5E56507E1 CRC64;

Query Match 76.1%; Score 35; DB 16; Length 345;  
Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8  
111111  
DB 217 YOHDLRAY 224

RESULT 9  
O919J4 PRELIMINARY; PRT: 392 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE CUN082 HYPOTHEICAL PROTEIN.  
GN CUN082.  
OS Culex nigripalpus baculovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae.  
OX NCBI\_TaxID=130556;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FLORIDA1997;  
RX MEDLINE=21488685; PubMed=11602755;  
RA Afonso C.L., Tlman E.R., Lu Z., Balinsky C.A., Moser B.A.,  
RA Becnel J.J., Rock D.L., Kutish G.F.;  
RT "Genome Sequence of a Baculovirus pathogenic for Culex nigripalpus."  
RL J. Virol. 75:11157-11165(2001).  
RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-FLORIDA1997;  
 RA Alfonso C.L., Tulman E.R., Lu Z., Balinsky C.A., Moser B.A.,  
 RA Becnel J.J., Rock D.L., Kutish G.F.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF403738; AAK94160.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 392 AA; 45338 MW; 49BE7B36DF3767E2 CRC64;

Query Match 76.1%; Score 35; DB 12; Length 392;  
 Best Local Similarity 75.0%; Pred. No. 56;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YOHDLR 8  
 |||||  
 Db 229 YOHDLRH 236

RESULT 10  
 ID Q9RS01 PRELIMINARY; PRT; 502 AA.  
 AC Q9RS01;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE SENSOR HISTIDINE KINASE.  
 GN DR2328.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Halt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 CC -1- SIMILARITY: TO PROKARYOTE SENSOR TRANSDUCTION PROTEINS.  
 DR EMBL: AE002064; AAF11875.1; -  
 DR TIGR: DR2328; -  
 DR InterPro: IPR004358; BCTRLSENSOR.  
 DR InterPro: IPR003660; HAMF.  
 DR InterPro: IPR003594; HATPase\_C.  
 DR InterPro: IPR003661; His\_KinA.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR Pfam: PF00672; HAMF; 1.  
 DR Pfam: PF02518; HATPase\_C; 1.  
 DR Pfam: PF00512; Signal; 1.  
 DR PRINTS: PR00344; BCTRLSENSOR.  
 DR SMART: SM00304; HAMF; 1.  
 DR SMART: SM00387; HATPase\_C; 1.  
 DR SMART: SM00388; HisK\_A; 1.  
 KW Complete proteome; Kinase; Phosphorylation; Sensory transduction;  
 KW Transferase.  
 SQ SEQUENCE 502 AA; 52988 MW; 295FD8FD050ECD0B CRC64;

Query Match 76.1%; Score 35; DB 16; Length 502;  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YOHDLR 6  
 |||||  
 Db 310 YOHDLR 315

RESULT 11  
 ID Q42136 PRELIMINARY; PRT; 167 AA.  
 AC Q42136;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TROPONIN C.  
 OS Lampetra japonica (Japanese lamprey) (Entosphenus japonicus).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hypocoartlia;  
 OC Petromyzontiformes; Petromyzontidae; Leptocheilichthys.  
 OX NCBI\_TaxID=94989;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98263050; PubMed=9604009;  
 RA Yusa H.J., Cox J.A., Takagi T.;  
 RT "Diversity of the troponin C genes during chordate evolution.";  
 RL J. Biochem. 123:1180-1190(1998).  
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
 DR EMBL: AB008555; BAA33282.1; -  
 DR HSSP: P10246; TRPF.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; EFhand; 4.  
 DR SMART: SM00054; EFh; 4.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_4.  
 KW Calcium-binding.  
 SQ SEQUENCE 167 AA; 18596 MW; C907CF1716AAAF20B CRC64;

Query Match 73.9%; Score 34; DB 13; Length 167;  
 Best Local Similarity 85.7%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 QHDLR 8  
 |||||  
 Db 10 QHDLR 16

RESULT 12  
 ID Q9U0R8 PRELIMINARY; PRT; 242 AA.  
 AC Q9U0R8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE DELTA-AMINOLEVULINIC ACID DEHYDRATASE (EC 4.2.1.24) (PORPHOBILINOGEN  
 DE SYNTHASE) (ALADH) (FRAGMENT).  
 GN HEM2.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20198189; PubMed=10731566;  
 RA Sato S., Tew S.I., Wilson R.J.M.;  
 RT "Impact of a plastid-beating endocytobiont on apicomplexan genomes.";  
 RL Int. J. Parasitol. 30:427-439(2000).  
 CC -1- CATALYTIC ACTIVITY: 2 5-AMINOLEVULINATE = PORPHOBILINOGEN + 2  
 H(2)O.  
 CC -1- PATHWAY: SECOND STEP IN PORPHYRIN AND HEME BIOSYNTHESIS.  
 CC -1- SUBUNIT: HOMOCYTRAMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ALADH FAMILY.  
 DR EMBL: AJ249365; CAB65342.1; -  
 DR HSSP: P15002; 18AE.  
 DR InterPro: IPR001731; D-ala-dehydratase.  
 DR Pfam: PF00490; ALAD; 2.  
 DR PRINTS: PR00144; DALDHDRATASE.  
 DR PROSITE: PD002304; D-ala-dehydratase; 1.  
 DR PROSITE: PS00169; D-ALA-DEHYDRATASE; 1.  
 KW Lyase; Porphyrin biosynthesis.  
 FT NON\_TER 1  
 FT NON\_TER 242

SQ SEQUENCE ~ 242 AA; 28534 MW; 7257D98516F82BBB CRC64;

Query Match 73.9%; Score 34; DB 5; Length 242;

Best Local Similarity 62.5%; Pred. No. 53;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8

|||||

Db 204 YOHAFMSY 211

RESULT 13

085470

ID 085470 PRELIMINARY; PRT; 316 AA.

AC 085470;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE SAGB (STREPTOLYSIN S ASSOCIATED ORF).

GN SAGB OR SPY0739.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1314;

RP SEQUENCE FROM N.A.

RC STRAIN=N2131.

RA Beall B., Nicholson M.L.;

RT "A Locus Required for Oxygen-Stable Hemolysis in Streptococcus

pyogenes.";

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;

RX MEDLINE=21192684; PubMed=11296296;

RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,

RA Primeaux C., Szate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,

RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,

RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;

RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).

DR EMBL: AF067649; AAC25178.1; -

DR EMBL: AE006526; AAC33685.1; -

DR InterPro: IPR000102; MAP1B\_neuraxin.

DR PROSITE: PS00230; MAP1B\_NEURAXIN; UNKNOWN\_1.

KW Complete proteome.

SQ SEQUENCE 316 AA; 35989 MW; D8926FB38E8E765 CRC64;

Query Match 73.9%; Score 34; DB 16; Length 316;

Best Local Similarity 75.0%; Pred. No. 70;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8

|||||

Db 204 YOHAFMSY 211

RESULT 14

09FVX2 PRELIMINARY; PRT; 481 AA.

AC 09FVX2;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HYPOTHETICAL 55.7 KDA PROTEIN.

GN F2P24.7.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,

RA Malt R., Konning C.M., Koo H., Fujii C.Y., Utterback T.R.,

RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome 1 BAC F2P24 genomic sequence.";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC078898; AAG29201.1; -

DR InterPro: IPR002885; PPR.

DR Pfam: PF01535; PPR; 10.

KW Hypothetical protein.

SQ SEQUENCE 481 AA; 55688 MW; 8F3E13B4D2FE3EDB CRC64;

Query Match 73.9%; Score 34; DB 10; Length 481;

Best Local Similarity 62.5%; Pred. No. 11e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8

|||||

Db 94 YEHVAVRY 101

RESULT 15

09HIE2 PRELIMINARY; PRT; 531 AA.

AC 09HIE2;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE HYPOTHETICAL PROTEIN TAI1398.

GN TAI1398.

OS Thermoplasma acidophilum.

OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;

OC Thermoplasma.

OX NCBI\_TaxID=2303;

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1728;

RX MEDLINE=20479972; PubMed=11029001;

RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,

RA Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;

RT "The genome sequence of the thermophilic scavenger Thermoplasma

acidophilum.";

RL Nature 407:508-513(2000).

DR EMBL: AL445067; CAC12518.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 531 AA; 60041 MW; 8F0C0BCD478307F6 CRC64;

Query Match 73.9%; Score 34; DB 17; Length 531;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YOHDLRAY 8

|||||

Db 271 YDNDLRAY 278

Search completed: August 15, 2002, 11:57:52  
Job time: 527 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: August 15, 2002, 11:39:44 ; Search time 73.67 Seconds  
(without alignments)  
12.062 Million cell updates/sec

Title: US-09-613-092a-5\_COPY\_3\_10

Perfect score: 51

Sequence: 1 SRVPMPTAW 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	100.0	15	20	AAV30351
2	42	82.4	34	22	ABB38515
3	42	82.4	34	22	ABB23664
4	42	82.4	34	22	AAW5133
5	42	82.4	34	22	AAW1668
6	42	82.4	34	22	AAW19262
7	42	82.4	34	22	AAW31962
8	42	82.4	81	22	ABW69150
9	39	76.5	1189	20	AAV15217
10	39	76.5	1189	20	AAV15218
11	38	74.5	10	16	AAW69771

12	38	74.5	15	21	AAW13080	Vascular endothel
13	38	74.5	27	22	ABB27332	Protein #9311 enco
14	38	74.5	53	20	AAW92467	Rat F-spondin FSP
15	38	74.5	58	22	AAW61606	Protonibacterium
16	38	74.5	72	22	AAW52332	Protonibacterium
17	38	74.5	76	22	ABW15813	Human nervous syst
18	38	74.5	84	22	AAW77743	Human colon cancer
19	38	74.5	102	22	ABW16216	Human nervous syst
20	38	74.5	106	22	ABW16994	Human nervous syst
21	38	74.5	128	22	AAW09709	Human gene 5 encod
22	38	74.5	157	22	AAW61821	Protonibacterium
23	38	74.5	168	22	AAW97743	H. annus SCIP-1 pr
24	38	74.5	182	21	AAW58354	Lung cancer associ
25	38	74.5	185	21	AAW14880	Arabidopsis thalia
26	38	74.5	343	22	AAW93099	Human protein sequ
27	38	74.5	436	21	AAW30975	Arabidopsis thalia
28	38	74.5	445	21	AAW5039	Arabidopsis thalia
29	38	74.5	454	21	AAW30974	Arabidopsis thalia
30	38	74.5	463	21	AAW52038	Arabidopsis thalia
31	38	74.5	525	22	AAW09697	Human gene 8 encod
32	38	74.5	568	14	AAW44243	Xenopus F-spondin
33	38	74.5	616	22	AAW39256	Human polypeptide
34	38	74.5	646	22	AAW41042	Human polypeptide
35	38	74.5	802	14	AAW44242	F-spondin (FP5-9) .
36	38	74.5	807	14	AAW44241	Human vasculat smo
37	38	74.5	807	20	AAW04262	Bovine vasculat sm
38	38	74.5	807	20	AAW04263	Rat vasculat smoot
39	38	74.5	807	20	AAW04264	Vascular endothel
40	38	74.5	807	21	AAW13072	Vascular endothel
41	38	74.5	807	21	AAW13073	Vascular endothel
42	38	74.5	807	21	AAW13074	Human polypeptide
43	38	74.5	807	22	AAW93799	Human ovarian canc
44	38	74.5	807	22	AAW74401	Novel human diagno
45	38	74.5	817	22	ABW30223	

#### ALIGNMENTS

RESULT 1  
AAV30351 standard; Peptide: 15 AA.  
AC AAV30351;  
XX  
XX 09-NOV-1999 (first entry)  
DE Epitope derived from pneumococcal surface adhesion A protein.  
XX  
XX Pneumococcal surface adhesion A protein; PsaA; monoclonal antibody;  
KW vaccine; Streptococcus pneumoniae infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX WO9945121-A1.  
XX  
XX 10-SEP-1999.  
PD  
XX  
XX 26-FEB-1999; 99WO-US04326.  
PF  
XX  
XX 02-MAR-1998; 98US-0076565.  
PR  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Aides EW, Carlone GW, Sampson JS, Tharpe JA, Westerink MAJ;  
PI Zeller JL;  
XX  
XX WPI; 1999-540849/45.  
PT New peptides corresponding to Streptococcus pneumoniae PsaA, used  
PT for treating or preventing Streptococcus pneumoniae infection in a  
PT subject  
XX

PS Claim 6; Page 43; 58bp; English.

XX AAY30351-54 represent immunogenic peptides which are derived from  
CC a pneumococcal surface adhesion A protein (PSA). The specification  
CC describes monoclonal antibodies which bind epitopes of the PSA protein  
CC (e.g. present sequence). The peptides can be used in vaccines to prevent  
CC Streptococcus pneumoniae infections. The antibodies of the invention  
CC can also be used to detect S. pneumoniae in a sample or individual.

SQ Sequence 15 AA;

Query Match 100.0%; Score 51; DB 20; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.08;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RVPWTAW 8  
|:|||||  
Db 3 RVPWTAW 10

RESULT 2

ID ABB38515 standard; Peptide; 34 AA.

AC ABB38515;

DT 04-FEB-2002 (first entry)

XX Peptide #6021 encoded by human foetal liver single exon probe.

KW Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001MO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

PS Claim 27; SEQ ID NO 31150; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC fetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 34 AA;

Query Match 82.4%; Score 42; DB 22; Length 34;

Best Local Similarity 71.4%; Pred. No. 4.6;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 RVPWTAW 8  
|:|||||  
Db 25 RVPWTAW 31

RESULT 3

ID ABB23664 standard; Protein; 34 AA.

AC ABB23664;

DT 23-JAN-2002 (first entry)

DE Protein #5663 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001MO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

PS Claim 15; SEQ ID NO 25434; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart (see

CC ABA21535-ABA41305). The present sequence is a protein encoded by one such

CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the

CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 34 AA;

Query Match 82.4%; Score 42; DB 22; Length 34;

Best Local Similarity 71.4%; Pred. No. 4.6;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 RVPWTAW 8  
|:|||||  
Db 25 RVPWTAW 31

```
RESULT 4
AAM59133
ID AAM59133 standard; Protein; 34 AA.
XX
AC AAM59133;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31238.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
DE Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 31238; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 34 AA;

Query Match 82.4%; Score 42; DB 22; Length 34;
Best Local Similarity 71.4%; Pred. No. 4.6;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RVPWTAW 8
DB 25 ripwtw 31

RESULT 5
AAM71668
ID AAM71668 standard; Protein; 34 AA.
XX
AC AAM71668;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31974.
XX
```

```
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
DE Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 31974; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 34 AA;

Query Match 82.4%; Score 42; DB 22; Length 34;
Best Local Similarity 71.4%; Pred. No. 4.6;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RVPWTAW 8
DB 25 ripwtw 31

RESULT 6
AAM19262
ID AAM19262 standard; Protein; 34 AA.
XX
AC AAM19262;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #5696 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
XX
```

XX	03-AUG-2000;	2000US-0632366.
PR	21-SEP-2000;	2000US-0234687.
PR	27-SEP-2000;	2000US-0236359.
PR	04-OCT-2000;	2000GB-0024263.
XX	(MOLE-) MOLECULAR DYNAMICS INC.	
XX	Penn SG,	Hanzel DK, Chen W, Rank DR;
XX	WPI;	2001-488901/53.
PT	Human genome-derived single exon nucleic acid probes useful for	
PT	analyzing gene expression in human cervical epithelial cells -	
XX	Claim 27; SEQ ID No 24088; 487bp; English.	
PS	The present invention relates to human single exon nucleic acid probes	
CC	(SNP: see A110068-A1128459). The present sequence is a peptide encoded	
CC	by one such probe. The SENPs are derived from human HeLa cells. The SENPs	
CC	can be used to produce a single exon microarray, which can be used for	
CC	measuring human gene expression in a sample derived from human cervical	
CC	epithelial cells. By measuring gene expression, the probes are therefore	
CC	useful in grading and/or staging of diseases of the cervix, notably	
CC	cervical cancer.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
SQ	Sequence	34 AA:
OY	2 RVPWFAM	8
Db	1	1
	25 ripwtcw	31
RESULT	7	
AAM31962	AAM31962 standard; Protein: 34 AA.	
ID	AAM31962	
AC	AAM31962;	
XX	17-OCT-2001 (first entry)	
DT	Peptide #999 encoded by probe for measuring placental gene expression.	
DE	Probe; microarray; human; placenta; antenatal diagnosis;	
XX	genetic disorder.	
KW	Homo sapiens.	
OS	WO200157272-A2.	
PN	09-AUG-2001.	
XX	30-JAN-2001; 2001WO-US00663.	
PF	04-FEB-2000; 2000US-0180312.	
XX	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX	(MOLE-) MOLECULAR DYNAMICS INC.	
PA	Penn SG,	Hanzel DK, Chen W, Rank DR;
XX	WPI;	2001-488901/53.
PT	Human genome-derived single exon nucleic acid probes useful for	
PT	analyzing gene expression in human cervical epithelial cells -	
XX	Claim 27; SEQ ID No 24088; 487bp; English.	
PS	The present invention relates to human single exon nucleic acid probes	
CC	(SNP: see A110068-A1128459). The present sequence is a peptide encoded	
CC	by one such probe. The SENPs are derived from human HeLa cells. The SENPs	
CC	can be used to produce a single exon microarray, which can be used for	
CC	measuring human gene expression in a sample derived from human cervical	
CC	epithelial cells. By measuring gene expression, the probes are therefore	
CC	useful in grading and/or staging of diseases of the cervix, notably	
CC	cervical cancer.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
SQ	Sequence	34 AA:
OY	2 RVPWFAM	8
Db	1	1
	25 ripwtcw	31

DR	WPI; 2001-488897/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for
PT	analyzing gene expression in human placenta -
XX	
PS	Claim 27; SEQ ID No 32231; 654bp; English.
XX	
CC	The present invention relates to single exon nucleic acid probes (SENP;
CC	see AAI13115-AA157546). The present sequence is a peptide encoded by one
CC	such probe. The probes are useful for producing a microarray for
CC	predicting, measuring and displaying gene expression in samples derived
CC	from human placenta. The probes are useful for antenatal diagnosis of
CC	human genetic disorders.
XX	
SQ	Sequence    34 AA:
Query Match	82.4%; Score 42; DB 22; Length 34;
Best Local Similarity	71.4%; Pred. No. 4.6;
Matches    5; Conservative	1; Mismatches    1; Indels    0; Gaps    0;
QY	2 RVPWTAW 8
	I : I I I I I
Db	25 ripwtw 31
RESULT	8
ABB69150	
ID	ABB69150 standard; Protein; 81 AA.
XX	
AC	ABB69150;
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 34242.
XX	
KM	Drosophila; developmental biology; cell signalling; insecticide;
KM	pharmaceutical.
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
Pf	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PERK ) PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-6556860/75.
N-PSDB; ABLI3253.	
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Disclosure: SEQ ID NO 34242; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC	sequences (ABLI01840-ABLI6175) and the encoded proteins
CC	(ABR57737-ABR72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO



CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 81 AA:

Query Match 82.4%, Score 42; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 3 VPWTAW 8  
|||||  
Db 58 vpwclaw 63

RESULT 9  
AA15217  
ID AA15217 standard; Protein: 1189 AA.

AC AA15217;

DT 09-NOV-1999 (first entry)

DE Human Hairless wildtype protein amino acid sequence.

XX alopecia; congenital alopecia; congenital atrichia;

KW androgenetic alopecia; alopecia areata; alopecia universalis;

KM wildtype; hair follicle.

XX Homo sapiens.

OS Homo sapiens.

PN MO9938965-A1.

XX 05-AUG-1999.

PD 29-JAN-1999; 99WO-US02128.

PR 29-JAN-1998; 98US-0073043.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

PA Christiano AM;

PI WPI: 1999-479184/40.

DR N-PSDB; AA206349.

XX Human hairless gene and protein, useful for identifying modulators

PT of hair growth

PS Claim 21; Fig 6; 127pp; English.

XX This is the amino acid sequence for the wildtype human hairless

CC protein.

CC The gene was discovered by genotyping a Pakistani kindred (comprising of

CC 4 affected males and 7 affected females) with an inherited form of

CC congenital alopecia universalis. The pedigree is strongly suggestive of

CC autosomal recessive inheritance.

CC The invention provides methods and sequences for the recombinant

CC production of wild-type human hairless, mutant human hairless and

CC wild-type human whn (winged-helix-nude) proteins, assays for screening

CC for blinding compounds, modulators and homologues, and animal models of

CC hairlessness. Human hairless conditions such as androgenetic alopecia

CC (male pattern baldness), alopecia areata, alopecia totalis,

CC congenital alopecia universalis, congenital alopecia and severe T-cell

CC immunodeficiency can be treated with compounds identified in the assays.

CC The methods are also useful for identifying compounds that can be used

CC to inhibit hair growth.

XX Sequence 1189 AA;

SQ

OY 3 VPWTAW 8  
|||||  
Db 274 vpwclaw 279

RESULT 10  
AA15218  
ID AA15218 standard; Protein: 1189 AA.

AC AA15218;

DT 09-NOV-1999 (first entry)

DE Human Hairless mutant amino acid sequence, Thr(1022)Ala.

XX alopecia; congenital alopecia; congenital atrichia;

KW androgenetic alopecia; alopecia areata; alopecia universalis;

KM mutant; hair follicle.

XX Homo sapiens.

OS Homo sapiens.

PN MO9938965-A1.

XX 05-AUG-1999.

PD 29-JAN-1999; 99WO-US02128.

PR 29-JAN-1998; 98US-0073043.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

PA Christiano AM;

PI WPI: 1999-479184/40.

DR Human hairless gene and protein, useful for identifying modulators

XX of hair growth

PS Claim 8; Page -: 127pp; English.

XX This is the amino acid sequence for the mutant human hairless

CC protein, which has Alanine substituted for Threonine at position

CC 1022.

CC The gene was discovered by genotyping a Pakistani kindred (comprising of

CC 4 affected males and 7 affected females) with an inherited form of

CC congenital alopecia universalis. The pedigree is strongly suggestive of

CC autosomal recessive inheritance.

CC The invention provides methods and sequences for the recombinant

CC production of wild-type human hairless, mutant human hairless and

CC wild-type human whn (winged-helix-nude) proteins, assays for screening

CC for blinding compounds, modulators and homologues, and animal models of

CC hairlessness. Human hairless conditions such as androgenetic alopecia

CC (male pattern baldness), alopecia areata, alopecia totalis,

CC congenital alopecia universalis, congenital alopecia and severe T-cell

CC immunodeficiency can be treated with compounds identified in the assays.

CC The methods are also useful for identifying compounds that can be used

CC to inhibit hair growth.

XX Sequence 1189 AA;

SQ

Query Match 76.5%; Score 39; DB 20; Length 1189;  
Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPWTAW 8  
|||||  
Query Match 76.5%; Score 39; DB 20; Length 1189;  
Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 274 vpwtsw 279

## RESULT 11

AAR69771 standard; peptide; 10 AA.

AAR69771;

13-OCT-1995 (first entry)

Thrombospondin peptide #6 for activating latent TGF-beta.

Thrombospondin type I repeat sequence; transforming growth factor-beta;

wound healing; fibrosis; endothelial cell proliferation.

Synthetic.

WO9505191-A.

23-FEB-1995.

12-AUG-1994; 94WO-US09193.

13-AUG-1993; 93US-0106120.

04-MAY-1994; 94US-0238169.

(UABR-) UAB RES FOUND.

Kruttsch HC, Murphy-Ullrich JE, Roberts DD, Schultz-Cherry S;

WPI; 1995-098579/13.

Stimulating or inhibiting transforming growth factor-beta by

contacting with thrombo-spondin or an activating enzyme - used

to enhance wound healing or prevent fibrosis

Claim 4; Page 62; 67pp; English.

The peptides AAR69766-79 are active peptide fragments based on the

second type I repeat sequence (amino acids 412-473) of thrombospondin

(sequence not given in the specification) which can be used to convert

latent transforming growth factor-beta (TGF-b) to active TGF-b.

Thrombospondin, or the activating peptides, can be used to stimulate

TGF-b for use in wound healing. Peptides (see AAR69780-90) which inhibit

the stimulation of TGF-b can be used to prevent fibrosis or block TGF-b

mediated endothelial cell proliferation.

Sequence 10 AA;

## Query Match

Best Local Similarity 74.5%; Score 38; DB 16; Length 10;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 PWTAW 8

4 pvtaw 8

21-DEC-2000 (first entry)

Vascular endothelial cell migration inhibitory protein SEQ ID 9.

Vascular endothelial cells migration; inhibition; abnormal angiogenesis;

solid tumour; inflammatory disease; arthritis; ocular angiogenic disease;

diabetic retinopathy.

XX OS Synthetic.

XX Key Location/Qualifiers

XX FT Modified-site 15 /note="C-terminal amide"

XX FT WO200047606-A1.

XX PD 17-AUG-2000.

XX PF 09-FEB-2000; 2000WO-JP00703.

XX PR 12-FEB-1999; 99JP-0033772.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Shibata K, Yamasaki M, Tsukuda E, Oda S, Miyamoto K;

XX DR WPI; 2000-549131/50.

XX PT New peptides inhibiting the migration of vascular endothelial cells for

XX PT treating diseases associated with abnormal angiogenesis including solid

XX PS tumors

XX PS Claim 4; Page 11; 58pp; Japanese.

XX CC This invention relates to peptides which inhibit the migration of

XX CC vascular endothelial cells. The peptides of the invention (sequences

XX CC AAB13075-B13085 including the present sequence) are fragments of the

XX CC human, bovine and rat sequences represented by sequences AAB13072-B13074.

XX CC The peptides are inhibitors of the migration of vascular endothelial

XX CC cells and are used for treating diseases associated with abnormal

XX CC angiogenesis including solid tumours, inflammatory diseases (e.g.

XX CC arthritis) and ocular angiogenic diseases (e.g. diabetic retinopathy).

Sequence 15 AA;

## Query Match

Best Local Similarity 74.5%; Score 38; DB 21; Length 15;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 PWTAW 8

4 pvtaw 7

23-JAN-2002 (first entry)

Protein #9331 encoded by probe for measuring heart cell gene expression.

Human: gene expression; heart; microarray; vascular system;

XX KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX KW congenital heart disease.

XX OS Homo sapiens.

XX PN WO200157274-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI: 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
PS Claim 15; SEQ ID No 29102; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA1305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 27 AA:

Query Match Best Local Similarity 74.5%; Score 38; DB 22; Length 27;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPWTAW 8  
: |||: |  
DB 20 lpwtsw 25

## RESULT 14

AAW92467 ID AAW92467 standard; Protein: 53 AA.

AC AAW92467;

DT 21-APR-1999 (first entry)

DE Rat F-spondin FSP-TSR-6 protein domain.

XX NAF-1; neuronal attachment factor-1; F-spondin analogue; treatment;  
XX spinal cord injury; peripheral nerves damage; neural cell adhesion;  
XX neurite extension; tumour cell metastasis; inhibitor; mobility; disease;  
XX endothelial cell proliferation; tumour neovascularisation; haemostasis;  
XX angiostatic agent; wound healing; diagnostic; neurotrophic; anticancer;  
XX antimetastatic; anti-angiogenic; antimalarial; rat; FSP; TSP; TSR;  
XX thrombospondin; thrombospondin type 1 repeat; FSP-TSR-6.

OS Rattus sp.

PN US5871969-A.

PD 16-FEB-1999.

PF 12-FEB-1997; 97US-0799173.

PR 12-FEB-1996; 96US-0011519.

XX 12-FEB-1997; 97US-0799173.

PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Dillon PJ, Hastings G;

XX WPI: 1999-166644/14.  
DR

XX Nucleic acid encoding human neuronal attachment factor-1 - used to  
PT e.g. treat spinal cord injuries, and inhibit tumor cell metastasis  
PT and neovascularization  
XX  
PS Disclosure: Fig 3; 29pp; English.

CC This sequence represents a rat F-spondin (FSP) thrombospondin (TSP) type  
CC 1 repeat domain, FSP-TSR-6 which is used in a method resulting in the  
CC isolation of a novel human neuronal attachment factor-1, NAF-1. NAF-1 is  
CC an analogue of rat F-spondin and is used to treat spinal cord injuries  
CC and damage to peripheral nerves (by promoting neural cell adhesion and  
CC neurite extension), to inhibit tumor cell metastasis (particularly in  
CC small cell and breast carcinoma) and endothelial cell proliferation,  
CC adhesion and mobility, to reduce tumor neovascularisation, as angiostatic  
CC agents for tumor cells, to promote wound healing and to modulate  
CC hemostasis. NAF-1 may also be used to identify treatments and diagnoses  
CC for human disease. Fragments of the nucleic acid that do not encode NAF-1  
CC peptides are useful as probes to isolate the NAF-1 gene, its allelic  
CC variants, full-length cDNA or related sequences, in chromosomal location  
CC by in situ hybridization or in Northern blotting, and as diagnostic  
CC probes or primers. The protein has neurotrophic, anticancer,  
CC antimetastatic, anti-angiogenic, antimalarial and modulates adhesion,  
CC proliferation and mobility of cells.

SQ Sequence 53 AA:

Query Match Best Local Similarity 74.5%; Score 38; DB 20; Length 53;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
: |||| |  
DB 5 pwtaw 9

## RESULT 15

AAU61606 ID AAU61606 standard; Protein: 58 AA.

AC AAU61606;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #22502.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-20841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI: 2001-616774/71.  
DR N-PSDB; AAS59618.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX

PS Example 1: SEQ ID No 22801; 1069pp; English.

XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 58 AA;

Query Match 74.58; Score 38; DB 22; Length 58;  
Best Local Similarity 71.4%; Pred. NO. 33;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 2 RVPWTAW 8  
| | | | |  
Db 20 repvtsw 26

Search completed: August 15, 2002, 11:49:01  
Job time: 557 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:44:59 ; Search time 27.13 Seconds  
(without alignments)  
7.203 Million cell updates/sec

Title: US-09-613-092a-5\_COPY\_3\_10  
Perfect score: 51  
Sequence: 1 SRVPMWTAW 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	76.5	984	4	US-09-287-354-2
2	39	76.5	1189	4	US-09-287-354-3
3	39	76.5	1189	4	US-09-287-354-4
4	38	74.5	53	2	US-08-799-173A-13
5	38	74.5	55	1	US-07-862-021B-20
6	38	74.5	55	5	PCT-US93-03164-20
7	38	74.5	568	1	US-07-862-021B-14
8	38	74.5	568	1	PCT-US93-03164-14
9	38	74.5	802	1	US-07-862-021B-12
10	38	74.5	802	1	US-08-313-288B-12
11	38	74.5	802	5	PCT-US93-03164-12
12	38	74.5	807	1	US-07-862-021B-10
13	38	74.5	807	1	US-08-313-288B-10
14	38	74.5	807	5	PCT-US93-03164-10
15	36	70.6	150	4	US-08-334-179A-6
16	36	70.6	150	4	US-08-334-179A-10
17	36	70.6	323	1	US-07-913-107-2
18	36	70.6	323	1	US-08-459-201-2
19	36	70.6	323	1	US-08-281-248-2
20	36	70.6	323	5	PCT-US93-06421-1
21	36	70.6	582	4	US-08-334-179A-4
22	36	70.6	870	2	US-08-732-192A-2
23	36	70.6	870	4	US-09-172-339-8
24	36	70.6	1038	4	US-08-334-179A-2
25	36	66.6	1038	4	US-08-334-179A-8
26	35	66.6	387	1	US-08-539-798-2
27	35	66.6	387	1	US-08-329-560-2

28	34	66.7	32	2	US-08-078-311-25	Sequence 25, Appl
29	34	66.7	32	2	US-08-460-402-25	Sequence 25, Appl
30	34	66.7	45	2	US-08-078-311-10	Sequence 10, Appl
31	34	66.7	45	2	US-08-460-402-10	Sequence 10, Appl
32	34	66.7	184	2	US-08-078-311-4	Sequence 4, Appl
33	34	66.7	184	2	US-08-460-402-4	Sequence 4, Appl
34	34	66.7	284	2	US-08-078-311-14	Sequence 14, Appl
35	34	66.7	284	2	US-08-460-402-14	Sequence 14, Appl
36	34	66.7	284	2	US-08-460-402-14	Sequence 14, Appl
37	34	66.7	284	2	US-08-460-402-14	Sequence 14, Appl
38	34	66.7	285	2	US-08-712-072C-5	Sequence 24, Appl
39	34	66.7	382	2	US-08-078-311-3	Sequence 3, Appl
40	34	66.7	382	2	US-08-460-402-3	Sequence 3, Appl
41	34	66.7	422	2	US-08-712-072C-2	Sequence 2, Appl
42	34	66.7	451	1	US-08-191-337-3	Sequence 3, Appl
43	34	66.7	482	1	US-08-184-337A-8	Sequence 8, Appl
44	34	66.7	482	5	PCT-US93-00670-8	Sequence 8, Appl
45	34	66.7	572	6	5200183-5	Patent No. 5200183

## ALIGNMENTS

RESULT 1  
US-09-287-354-2  
; Sequence 2, Application US/09287354  
; Patent No. 6348348  
; GENERAL INFORMATION:  
; APPLICANT: THOMPSON, Catherine C.  
; TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN  
; FILE REFERENCE: Thompson-20263/0243435  
; CURRENT APPLICATION NUMBER: US/09/287,354  
; CURRENT FILING DATE: 1999-04-07  
; EARLIER APPLICATION NUMBER: US 60/080,888  
; EARLIER FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 984  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-287-354-2

Query Match 76.5%; Score 39; DB 4; Length 984;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 VPWTAW 8  
| | | | |  
DB 69 VPWTAW 74

RESULT 2  
US-09-287-354-3  
; Sequence 3, Application US/09287354  
; Patent No. 6348348  
; GENERAL INFORMATION:  
; APPLICANT: THOMPSON, Catherine C.  
; TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN  
; FILE REFERENCE: Thompson-20263/0243435  
; CURRENT APPLICATION NUMBER: US/09/287,354  
; CURRENT FILING DATE: 1999-04-07  
; EARLIER APPLICATION NUMBER: US 60/080,888  
; EARLIER FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1189  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-287-354-3

Query Match 76.5%; Score 39; DB 4; Length 1189;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPWTAW 8  
||||:|  
Db 274 VPWTAW 279

RESULT 3  
US-09-287-354-4  
; Sequence 4, Application US/09287354  
; Patent No. 6348348  
; GENERAL INFORMATION:  
; APPLICANT: THOMPSON, Catherine C.  
; TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN  
; FILE REFERENCE: Thompson-20263/0243435  
; CURRENT APPLICATION NUMBER: US/09/287,354  
; CURRENT FILING DATE: 1999-04-07  
; EARLIER APPLICATION NUMBER: US 60/080,888  
; EARLIER FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1189  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-287-354-4

Query Match 76.5%; Score 39; DB 4; Length 1189;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPWTAW 8  
||||:|  
Db 274 VPWTAW 279

RESULT 4  
US-08-799-173A-13  
; Sequence 13, Application US/08799173A  
; Patent No. 5871969  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, GREGG,  
; APPLICANT: PATRICK J. DILLON  
; TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
; STREET: 9410 KEY WEST AVENUE  
; CITY: ROCKVILLE  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/799,173A  
; FILING DATE: 11-FEB-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROOKES, ANDERS A.  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF226  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-799-173A-13

Query Match 74.5%; Score 38; DB 2; Length 53;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
Db 5 PWTAW 9

RESULT 5  
US-07-862-021B-20  
; Sequence 20, Application US/07862021B  
; Patent No. 5279966  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Klar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,021B  
; FILING DATE: 19920405  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 55 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-862-021B-20

Query Match 74.5%; Score 38; DB 1; Length 55;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
Db 7 PWTAW 11

RESULT 6  
PCT-US93-03164-20

; Sequence 20, Application PC/TUS9303164  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Klar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/03164  
; FILING DATE: 19930402  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 55 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US93-03164-20  
  
Query Match 74.5%; Score 38; DB 5; Length 55;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 4 PWTAW 8  
Db 7 PWTAW 11  
  
RESULT 7  
US-07-862-021B-14  
; Sequence 14, Application US/07862021B  
; Patent No. 5279966  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Klar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/07/862,021B  
; APPLICATION NUMBER: 19920405  
; FILING DATE: 19920405  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 568 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-862-021B-14  
  
Query Match 74.5%; Score 38; DB 1; Length 568;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 4 PWTAW 8  
Db 520 PWTAW 524  
  
RESULT 8  
PCT-US93-03164-14  
; Sequence 14, Application PC/TUS9303164  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Klar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/03164  
; FILING DATE: 19930402  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 568 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear

```
; MOLECULE TYPE: protein
PCT-US93-03164-14

Query Match          74.5%; Score 38; DB 5; Length 568;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PWTAW 8
   |||||
Db 520 PWTAW 524

RESULT 9
US-07-862-021B-12
; Sequence 12, Application US/07862021B
; Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-021B-12

Query Match          74.5%; Score 38; DB 1; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PWTAW 8
   |||||
Db 754 PWTAW 758

RESULT 10
US-08-313-288B-12
; Sequence 12, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
```

```
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-313-288B-12

Query Match          74.5%; Score 38; DB 1; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PWTAW 8
   |||||
Db 754 PWTAW 758

RESULT 11
PCT-US93-03164-12
; Sequence 12, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
```



REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-03164-12

Query Match 74.5%; Score 38; DB 1; Length 807;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
DB 754 PWTAW 758

RESULT 12  
US-07-862-021B-10  
Sequence 10, Application US/07862021B  
Patent No. 5279966  
GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M  
APPLICANT: Klar, Avihu  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,021B  
FILING DATE: 19920405  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 807 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-862-021B-10

Query Match 74.5%; Score 38; DB 1; Length 807;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 4 PWTAW 8  
|||||

DB 759 PWTAW 763

RESULT 13  
US-08-313-288B-10  
Sequence 10, Application US/08313288B  
Patent No. 5750502  
GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M. and Avihu Klar  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,288B  
FILING DATE: January 5, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 807 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-313-288B-10

Query Match 74.5%; Score 38; DB 1; Length 807;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
DB 759 PWTAW 763

RESULT 14  
PCT-US93-03164-10  
Sequence 10, Application PC/TUS9303164  
GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M  
APPLICANT: Klar, Avihu  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03164  
FILING DATE: 19930402  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SFO ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 807 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-03164-10

Query Match 74.5%; Score 38; DB 5; Length 807;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PWTAW 8  
|||||  
Db 759 PWTAW 763

RESULT 15  
US-08-334-179A-6  
Sequence 6, Application US/08334179A  
Patent No. 6306622  
GENERAL INFORMATION:  
APPLICANT: ROSENBAUM, JAN S.  
APPLICANT: NOHNO, TSUTOMU  
TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THE PROCTER AND GAMBLE COMPANY  
STREET: 11810 EAST MIAMI RIVER ROAD  
CITY: ROSS  
STATE: OH  
COUNTRY: US  
ZIP: 45061  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.30, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/334,179A  
FILING DATE: 04-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CORSTANJE, BRAHM J.  
REGISTRATION NUMBER: 34,804  
REFERENCE/DOCKET NUMBER: 5473  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 513-627-2858  
TELEFAX: 513-627-0260  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 150 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-334-179A-6

Query Match 70.6%; Score 36; DB 4; Length 150;  
Best Local Similarity 71.4%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RVPWTAW 8  
|||||  
Db 10 RVPWLPW 16

Search completed: August 15, 2002, 11:49:35  
Job time: 276 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:46:14 ; Search time 35.29 Seconds  
(without alignments)  
21.783 Million cell updates/sec

Title: US-09-613-092a-5\_COPY\_3\_10

Perfect score: 51

Sequence: 1 SRVPTAW 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapect 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

PIR-71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	76.5	335	2	H72420
2	39	76.5	497	2	C49850
3	38	74.5	162	2	T48152
4	38	74.5	441	2	G83212
5	38	74.5	454	2	G86284
6	38	74.5	483	2	T06712
7	38	74.5	489	2	T06715
8	38	74.5	616	2	I38155
9	38	74.5	773	2	G84554
10	38	74.5	803	2	A47723
11	38	74.5	807	2	A38152
12	38	74.5	870	1	DPECN
13	38	74.5	870	2	G90755
14	38	74.5	870	2	E85619
15	37	72.5	309	2	E86566
16	37	72.5	490	2	T06714
17	37	72.5	2290	1	GNNYE
18	37	72.5	2292	2	S75654
19	36	70.6	291	2	S75654
20	36	70.6	316	2	H95975
21	36	70.6	323	2	A39877
22	36	70.6	328	2	T37734
23	36	70.6	351	2	S63197
24	36	70.6	397	2	A49349
25	36	70.6	437	2	T47831
26	36	70.6	490	2	T06711
27	36	70.6	541	2	A87595
28	36	70.6	541	2	A33598
29	36	70.6	823	2	T02812

30	36	70.6	851	2	T47495	hypothetical prote
31	36	70.6	1038	2	I38935	bone morphogenetic
32	36	70.6	1038	2	JC5527	bone morphogenetic
33	36	70.6	1524	2	G84721	hypothetical prote
34	36	70.6	1750	2	G84649	hypothetical prote
35	35	68.6	122	2	J00150	hypothetical 13k p
36	35	68.6	133	2	AF0447	probable phage-rel
37	35	68.6	182	2	T24511	hypothetical prote
38	35	68.6	262	2	T52310	L113 protein [limp
39	35	68.6	277	2	A46241	interferon respons
40	35	68.6	280	2	C70023	retinol dehydrogen
41	35	68.6	303	1	QOECRT	L(+)-tartrate dehy
42	35	68.6	303	2	H91121	L-tartrate dehydra
43	35	68.6	303	2	G85966	L-tartrate dehydra
44	35	68.6	312	2	C71243	hypothetical prote
45	35	68.6	313	2	C75208	sugar transport sy

#### ALIGNMENTS

RESULT 1  
H72420  
oligopeptide ABC transporter, permease protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72420  
R:Neilsen, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316  
A:Accession: H72420  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-335 <ARN>  
A:Cross-references: GB:AE001694; GB:AE000512; NID:g4980558; PIDN:AMD35166.1; PID:g498  
C:Genetics:  
A:Experimental source: strain MSB8  
C:Superfamily: oligopeptide permease protein opps  
Query Match 76.5%; Score 39; DB 2; Length 335;  
Best Local Similarity 83.3%; Pred. No. 42;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 VPWTAW 8  
DB 109 LPWTAW 114  
RESULT 2  
C49850  
proteolipophyllide reductase (EC 1.3.1.33) 52.5k chain - Rhodobacter capsulatus  
N:Alternate names: chlorin reductase subunit bchy  
C:Species: Rhodobacter capsulatus  
C:Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 17-Mar-2000  
C:Accession: C49850; S17824  
R:Burke, D.H.; Alberti, M.; Hearst, J.E.  
J. Bacteriol. 175, 2407-2413, 1993  
A:Title: The Rhodobacter capsulatus chlorin reductase-encoding locus, bchy, consists  
A:Reference number: A49850; MUID:93224464  
A:Accession: C49850  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-497 <BUR>  
A:Cross-references: EMBL:Z11165; NID:g46097; PIDN:CAA77549.1; PID:g46132  
A:Experimental source: strain SB1003  
A:Note: Sequence extracted from NCBI backbone (NCBIN:129230, NCBIPI:129233)  
C:Genetics:  
A:Gene: bchy

C:Superfamily: protochlorophyllide reductase chain chlB  
C:Keywords: oxidoreductase; transmembrane protein

Query Match 76.5% Score 39; DB 2; Length 497;  
Best Local Similarity 71.4%; Pred. No. 61;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 2 RVPTAW 8  
Db 356 RRPMSAW 362

RESULT 3  
T48152  
hypothetical protein T1008.10 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 21-Jul-2000  
C:Accession: T48152  
R:Bevan, M.; Pohl, T.; Welzenegger, T.; Bancroft, I.; Mewes, H.W.; Lemcke, K.; Mayer, K.  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24486  
A:Accession: T48152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-162 <BEV>  
A:Cross-references: EMBL:AL161746  
A:Experimental source: cultivar Columbia; BAC clone T1008  
C:Genetics:  
A:Map position: 5  
A:Introns: 40/2; 138/3  
A:Note: T1008.10  
C:Superfamily: conserved hypothetical protein ybhB

Query Match 74.5% Score 38; DB 2; Length 162;  
Best Local Similarity 83.3%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 VPWTAW 8  
Db 67 VPWTVM 72

RESULT 4  
G83212  
hypothetical protein PA3464 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83212  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: AB2950; MUID:20437337  
A:Accession: G83212  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-441 <STO>  
A:Cross-references: GB:AE004767; GB:AE004091; NID:99949604; PIDN:AA606852.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3464

Query Match 74.5% Score 38; DB 2; Length 441;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
Db 144 PWTAW 148

RESULT 5  
G86284  
F911.4 protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
C:Accession: G86284  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: G86284  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <STO>

A:Cross-references: GB:AE005172; NID:95103809; PIDN:AAD39639.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 74.5% Score 38; DB 2; Length 454;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
Db 32 PWTAW 36

RESULT 6  
T06712  
probable cytochrome P450 T29H1.180 - Arabidopsis thaliana  
N:Alternate names: protein T29H1.180  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 04-Mar-2000  
C:Accession: T06712  
R:Queller, F.; Choisme, N.; Robert, C.; Brotlier, P.; Wincker, P.; Cattoilco, L.; Art  
submitted to the Protein Sequence Database, April 1999  
A:Reference number: Z15793  
A:Accession: T06712  
A:Molecule type: DNA  
A:Residues: 1-483 <QUE>  
A:Cross-references: EMBL:AL049659  
A:Experimental source: cultivar Columbia; BAC clone T29H11  
C:Genetics:  
A:Map position: 3  
A:Introns: 284/3  
A:Note: T29H1.180  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
F:286-451/Domain: cytochrome P450 homology <CT>  
F:429/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 74.5% Score 38; DB 2; Length 483;  
Best Local Similarity 75.0%; Pred. No. 85;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRVPWTAW 8  
Db 211 SYVPWLAW 218

RESULT 7

T06715  
Probable cytochrome P450 T29H1.210 - Arabidopsis thaliana  
N:Alternate names: protein T29H1.210  
M:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 04-Mar-2000  
C:Accession: T06715  
R:Querliar, F.; Choisme, N.; Robert, C.; Brotier, P.; Wincker, P.; Catolico, L.; Artig  
submitted to the Protein Sequence Database, April 1999  
A:Reference number: 215793  
A:Accession: T06715  
A:Molecule type: DNA  
A:Residues: 1-489 <QUE>  
A:Cross-references: EMBL:AL049659  
A:Experimental source: cultivar Columbia; BAC clone T29H1  
C:Genetics:  
A:Map position: 3  
A:Introns: 286/3  
A:Note: T29H1.210  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C:Keywords: Chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F:288-453/Domain: cytochrome P450 homology <CYP>  
F:431/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 74.5%; Score 38; DB 2; Length 489;  
Best Local Similarity 75.0%; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRVPTAW 8  
|||  
Db 213 SYVPLAW 220

RESULT 8  
138155  
DNA-binding regulatory factor X5 - human  
N:Alternate names: protein DKFZp586K091.1; RFX5 protein  
C:Species: Homo sapiens (man)  
C:Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 21-Jul-2000  
C:Accession: 138155; T08770  
R:Steimle, V.; Durand, B.; Barras, E.; Zufferey, M.; Hadam, M.R.; Mach, B.; Relth, W.  
Genes Dev. 9, 1021-1032, 1995  
A:Title: A novel DNA-binding regulatory factor is mutated in primary MHC class II defect  
A:Reference number: 138155; MUID:95262896  
A:Accession: 138155  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-616 <RES>  
A:Cross-references: EMBL:X85786; NID:q840788; PIDN:CMA59771.1; PID:q840789  
A:Experimental source: B cell line RAJI  
R:Ottewaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: 216474  
A:Accession: T08770  
A:Molecule type: mRNA  
A:Residues: 183-616 <OTU>  
A:Cross-references: EMBL:AL050135  
A:Experimental source: adult uterus; clone DKFZp586K091  
C:Genetics:  
A:Gene: GDB:RFX5  
A:Cross-references: GDB:6288464; OMIM:601863  
A:Map position: 19p13.3-19p13.3  
A:Note: DKFZp586K091.1

Query Match 74.5%; Score 38; DB 2; Length 616;  
Best Local Similarity 62.5%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRVPTAW 8  
|||  
Db 496 SRVPTAW 503

RESULT 9  
C84554  
Hypothetical protein At2g17610 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84554  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Yankin, S.E.; Unayam, L.; Tallon,  
euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: C84554  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-773 <STO>  
A:Cross-references: GB:AE002093; NID:q4926870; PIDN:AD32950.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g17610  
A:Map position: 2

Query Match 74.5%; Score 38; DB 2; Length 773;  
Best Local Similarity 57.1%; Pred. No. 1.3e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 RVPTAW 8  
::|  
Db 290 KIPVWAW 296

RESULT 10  
A47723  
F-spondin precursor - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 17-Nov-2000  
C:Accession: A47723  
R:Ruiz I Altaba, A.; Cox, C.; Jessell, T.M.; Klar, A.  
Proc. Natl. Acad. Sci. U.S.A. 90, 8268-8272, 1993  
A:Title: Ectopic neural expression of a floor plate marker in frog embryos injected w  
A:Reference number: A47723; MUID:93376785  
A:Accession: A47723  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-803 <RUT>  
A:Cross-references: GB:L09123; NID:q409244; PIDN:AA19105.1; PID:q409245  
C:Superfamily: F-spondin; thrombospondin type 1 repeat homology <THR2>  
F:435-489/Domain: thrombospondin type 1 repeat homology <THR2>  
F:607-663/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 74.5%; Score 38; DB 2; Length 803;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
||||  
Db 755 PWTAW 759

RESULT 11  
A38152  
F-spondin - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Nov-2000  
C:Accession: A38152  
R:Klar, A.; Baldassare, M.; Jessell, T.M.  
Cell 69, 95-110, 1992  
A:Title: F-spondin: a gene expressed at high levels in the floor plate encodes a secr  
A:Reference number: A38152; MUID:92208952  
A:Accession: A38152

A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-807 <RLA>  
A:Cross-references: GB:M88469; NID:g204176; PIDN:AAA41174.1; PID:g204177  
A:Experimental source: embryo floor plate  
A>Note: sequence extracted from NCBI backbone (NCBIN:90877, NCBI:P:90878)  
C:Superfamily: F-spondin; thrombospondin type 1 repeat homology  
F:441-495/Domain: thrombospondin type 1 repeat homology <THR2>  
F:500-555/Domain: thrombospondin type 1 repeat homology <THR3>  
F:557-611/Domain: thrombospondin type 1 repeat homology <THR4>  
F:613-666/Domain: thrombospondin type 1 repeat homology <THR5>  
F:667-721/Domain: thrombospondin type 1 repeat homology <THR5>  
F:753-807/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 74.5%; Score 38; DB 2; Length 807;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWM 8  
|||||  
Db 759 PWM 763

RESULT 12  
DPECN  
membrane alanyl aminopeptidase (EC 3.4.11.2) - Escherichia coli  
N:Alternate names: alpha-aminocylpeptide hydrolase; aminopeptidase N; microsomal aminop  
C:Species: Escherichia coli  
C>Date: 30-Jun-1988 #sequence\_revision 26-Sep-1997 #text\_change 01-Feb-2002  
C:Accession: C64833; A27164; A91561; B91163; I57748; A25058; A29045  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Co  
.A.: Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; M0ID:97426617  
A:Accession: C64833  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-870 <BLAT>  
A:Cross-references: GB:AE000195; GB:U00096; NID:g1787156; PIDN:NA074018.1; PID:g1787163;  
A:Experimental source: strain K-12, substrain MG1655  
R:Fogolino, M.; Gharbi, S.; Lazdunski, A.  
Gene 49, 303-309, 1986  
A:Title: Nucleotide sequence of the pepN gene encoding aminopeptidase N of Escherichia c  
A:Reference number: A27164; M0ID:87192020  
A:Accession: A27164  
A:Molecule type: DNA  
A:Residues: 1-870 <ROG>  
A:Cross-references: GB:M15676; NID:g147143; PIDN:AAA24318.1; PID:g147144  
R:McCaman, M.T.; Gabe, J.D.  
Gene 48, 145-153, 1986  
A:Title: The nucleotide sequence of the pepN gene and its over-expression in Escherichia  
A:Reference number: A91561; M0ID:87163509  
A:Accession: A91561  
A:Molecule type: DNA  
A:Residues: 1-642 'R', 644-870 <MCC>  
R:Bally, M.; Foglino, M.; Bruschl, M.; Murgier, M.; Lazdunski, A.  
Eur. J. Biochem. 155, 565-569, 1986  
A:Title: Nucleotide sequence of the promoter and amino-terminal encoding region of the B  
A:Reference number: A91163; M0ID:86164315  
A:Accession: A91163  
A:Molecule type: DNA  
A:Residues: 1-176 <BA1>  
A:Cross-references: GB:X03709; NID:g42354; PIDN:CAA27336.1; PID:g42356  
A:Accession: B91163  
A:Molecule type: protein  
A:Residues: 1-21 <BA2>  
R:McCaman, M.T.; Gabe, J.D.  
Mol. Gen. Genet. 204, 148-152, 1986  
A:Title: Sequence of the promoter and 5' coding region of pepN, and the amino-terminus C  
A:Reference number: I57748; M0ID:86310300  
A:Accession: I57748

A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-75, 'D', 77-242 <RES>  
A:Cross-references: EMBL:X04020; NID:g42352; PIDN:CAA27647.1; PID:g42353  
A:Comment: This enzyme is active in the cytoplasm and is probably bound to the inner  
C:Comment: This enzyme hydrolyzes L-amino acid beta-naphthylamides.  
C:Genetics:  
A:Gene: pepN  
A:Map position: 21 min  
C:Function:  
A:Description: releases N-terminal amino acid residues from polypeptides and proteins  
A:Note: requires zinc for proteolytic activity  
C:Superfamily: microsomal aminopeptidase  
C:Keywords: aminopeptidase; membrane protein; metalloproteinase; zinc  
F:297,301,320/Binding site: zinc, catalytic (His, His, Glu) #status predicted  
F:381/Active site: Tyr #status predicted

Query Match 74.5%; Score 38; DB 1; Length 870;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWM 8  
|||||  
Db 70 PWM 74

RESULT 13  
G90755  
aminopeptidase N [imported] - Escherichia coli (strain 0157:H7, substrain RIMD 050995  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: G90755  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g  
A:Reference number: A9629; M0ID:21156231; PMID:11258796  
A:Accession: G90755  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-870 <NAV>  
A:Cross-references: GB:BA000007; PIDN:BA034438.1; PID:g13360474; GSPDB:GN00154  
A:Experimental source: strain 0157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECS1015  
C:Superfamily: microsomal aminopeptidase

Query Match 74.5%; Score 38; DB 2; Length 870;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWM 8  
|||||  
Db 70 PWM 74

RESULT 14  
E85619  
aminopeptidase N [imported] - Escherichia coli (strain 0157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: E85619  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grobeck, E.J.; Davis, N.W.; Llim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.  
A:Reference number: A85480; M0ID:21074935; PMID:11206951  
A:Accession: E85619  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-870 <STO>



A:Cross-references: GB:AE005174; NID:g12514107; PIDN:AG55417.1; GSPDB:GN00145; UMGCP:212  
 A:Experimental source: strain 0157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: pepN  
 C:Superfamily: microsomal aminopeptidase

Query Match 74.5%; Score 38; DB 2; Length 870;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PWTAW 8  
 |||||  
 DB 70 PWTAW 74

RESULT 15  
 E96566  
 F6D8.20 [Imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E96566  
 R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;  
 ansen, N.F.; Hughes, B.; Hulzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: E96566  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-309 <STO>  
 A:Cross-references: GB:AE005173; NID:g5903047; PIDN:AAD55606.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F6D8.20  
 A:Map position: 1

Query Match 72.5%; Score 37; DB 2; Length 309;  
 Best Local Similarity 83.3%; Pred. No. 80;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPWTAW 8  
 |||||  
 DB 29 VPWTAW 34

Search completed: August 15, 2002, 11:50:17  
 Job time: 243 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:49:40 ; Search time 18.61 seconds

(without alignments)  
16.645 Million cell updates/sec

Title: US-09-613-092a-5\_COPY\_3\_10

Perfect score: 51

Sequence: 1 SRVPMWAW 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	39	76.5	497	1	BCHY_RHOCA
2	39	76.5	1189	1	HAIR_HUMAN
3	38	74.5	483	1	C71M_ARATH
4	38	74.5	489	1	C71Q_ARATH
5	38	74.5	616	1	REF5_HUMAN
6	38	74.5	803	1	ESPO_XENLA
7	38	74.5	807	1	FSPO_RAT
8	38	74.5	869	1	AMPN_ECOLI
9	37	72.5	490	1	C71P_ARATH
10	37	72.5	2290	1	POLG_EMCV
11	36	70.6	316	1	EXOV_RHIME
12	36	70.6	323	1	AMFR_HUMAN
13	36	70.6	351	1	YXN1_YEAST
14	36	70.6	395	1	SYT8_MOUSE
15	36	70.6	650	1	C71M_ARATH
16	36	70.6	498	1	GUN3_FIBSU
17	36	70.6	1038	1	BMR2_MOUSE
18	36	70.6	1038	1	BMR2_MOUSE
19	35	68.6	280	1	YUSZ_BACSU
20	35	68.6	303	1	TTDA_ECOLI
21	35	68.6	708	1	CAO2_CANTR
22	35	68.6	1074	1	SMSA_HUMAN
23	35	68.6	1077	1	SMSA_MOUSE
24	35	68.6	1120	1	RPMW_SCHPO
25	35	68.6	1180	1	TRK2_MOUSE
26	34	66.7	119	1	VFX_STIVAI
27	34	66.7	119	1	VFX_STIVAI
28	34	66.7	119	1	VFX_STIVAI
29	34	66.7	184	1	MPL_MPLV
30	34	66.7	280	1	SOL3_YEAST
31	34	66.7	350	1	YN04_YEAST
32	34	66.7	426	1	PIRC_THEAO
33	34	66.7	470	1	MTR_NEUCR

34	34	66.7	541	1	BRL1_SCHCO	P78741 schizoplyll
35	34	66.7	581	1	CHLE_RABIT	P21927 oryctolaus
36	34	66.7	595	1	SNX9_HUMAN	O955x1 homo sapien
37	34	66.7	602	1	CHLE_HUMAN	P06276 homo sapien
38	34	66.7	603	1	CHLE_MOUSE	O08311 mus musculu
39	34	66.7	625	1	TPOR_MOUSE	O08351 mus musculu
40	34	66.7	775	1	MGD1_MOUSE	O9qy46 mus musculu
41	34	66.7	775	1	MGD1_RAT	O9q573 rattus norv
42	34	66.7	778	1	MGD1_HUMAN	O955v3 homo sapien
43	34	66.7	990	1	TNP7_ECOLI	P13694 escherichia
44	34	66.7	1385	1	RRPO_P1AV	O07518 plantago as
45	34	66.7	1473	1	NAL1_HUMAN	O9c000 homo sapien

## ALIGNMENTS

RESULT	ID	BCHY_RHOCA	STANDARD	PRT	497 AA
AC	P26178:				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Chlorophyllide reductase 52.5 kDa chain (EC 1.18.1.-) (Chlorin reductase).				
GN	BCHY.				
OS	Rhodobacter capsulatus (Rhodospseudomonas capsulata).				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.				
OX	NCBI_TaxID=1061;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SB1003;				
RX	MEDLINE=93224464; PubMed=8468299;				
RA	Burke D.H., Alberti M., Hearst J.E.;				
RT	"The Rhodobacter capsulatus chlorin reductase-encoding locus, bcha, consists of three genes, bchy, bchy, and bchz.";				
RL	J. Bacteriol. 175:2407-2413(1993).				
CC	-1- FUNCTION: CONVERTS CHLOROPHYLLS (CHL) INTO BACTERIOCHLOROPHYLLS (BCHL) BY REDUCING RING B OF THE TETRAPIROLE.				
CC	-1- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.				
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).				
CC	-1- SIMILARITY: BELONGS TO THE BCHN / CHLN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL, Z11165; CAA7549.1; -.				
DR	PIR, S17824; S17824.				
DR	PIR, C49850; C49850.				
KW	Oxidoreductase; Photosynthesis; Bacteriochlorophyll biosynthesis;				
KW	Transmembrane.				
FT	TRANSMEM 65 82				POTENTIAL.
FT	TRANSMEM 126 142				POTENTIAL.
FT	TRANSMEM 216 233				POTENTIAL.
SQ	SEQUENCE 497 AA; 52618 MW; 99369F0EA91A4209 CRC64;				

Query Match 76.5%; Score 39; DB 1; Length 497;  
Best Local Similarity 71.4%; Pred. No. 21;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RVPMWAW 8  
DB 356 RTPWSAW 362

RESULT 2  
HAIR\_HUMAN STANDARD: PRT: 1189 AA.  
AC 043593; Q9NPE1:  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Hairless protein.  
GN HR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM), AND VARIANT ALUNC ALA-1022.  
RX MEDLINE=9811413; PubMed=9445480;  
RA Ahmad W., ul Haque M.F., Brancolini V., Tsou H.C., ul Haque S.,  
Lam H., Alta V.M., Owen J., Deblaquiere M., Frank J.,  
Ra Ceerhalml-Friedman P.B., Leask A., McGrath J.A., Peacocke M.,  
Ahmad M., Olt J., Cristiano A.M.;  
RT "Alopecia universalis associated with a mutation in the human hairless  
gene";  
RL Science 279:720-724(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (LONG ISOFORM), REVISIONS TO 572 AND 774, AND  
RP TISSUE SPECIFICITY.  
RC TISSUE-Peripheral blood leukocytes, and Skin fibroblast;  
RX MEDLINE=99162400; PubMed=10051399;  
RA Ahmad W., Zlotogorski A., Panteleyev A.A., Lam H., Ahmad M.,  
ul Haque M.F., Abdallah H.M., Dragan L., Cristiano A.M.;  
RT "Genomic organization of the human hairless gene (HR) and  
RT identification of a mutation underlying congenital atrichia in an Arab  
RT Palestinian family";  
RL Genomics 56:141-148(1999).  
RN [3]  
RP SEQUENCE FROM N.A., VARIANT ALUNC ASP-1136, TISSUE SPECIFICITY, AND  
RP ALTERNATIVE SPLICING.  
RC TISSUE-Peripheral blood leukocytes, Brain, and Fetal brain;  
RX MEDLINE=98409496; PubMed=9736769;  
RA Cichon S., Anker M., Vogt I.R., Rohleder H., Putzstuck M., Hillemer A.,  
Farooq S.A., Al-Dhafiri K.S., Ahmad M., Haque S., Rletschel M.,  
Ra Propping P., Kruse R., Noethen M.M.;  
RT "Cloning, genomic organization, alternative transcripts and mutational  
RT analysis of the gene responsible for autosomal recessive universal  
RT congenital alopecia";  
RL Hum. Mol. Genet. 7:1671-1679(1998).  
RN [4]  
RP VARIANT APL GLN-620.  
RX MEDLINE=98431781; PubMed=9758627;  
RA Ahmad W., Irvine A.D., Lam H., Buckley C., Bingham E.A.,  
Panteleyev A.A., Ahmad M., McGrath J.A., Cristiano A.M.;  
RT "A missense mutation in the zinc-finger domain of the human hairless  
RT gene underlies congenital atrichia in a family of Irish travellers";  
RL Am. J. Hum. Genet. 63:984-991(1998).  
RN [5]  
RP FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO  
RP REGULATE ONE OF THE PHASES OF HAIR GROWTH.  
CC -1- SUBCELLULAR LOCATION: Nucleat.  
CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms: a long form (shown  
CC here) and a short form; are produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Strongest expression of both isoforms is  
CC in the small intestine, weaker expression in brain and colon, and  
CC trace expression is found in liver, pancreas, spleen, thymus,  
CC stomach, salivary gland, appendix and trachea. Long isoform is  
CC always the most abundant. Long isoform is exclusively expressed at  
CC low levels in kidney and testis and short isoform exclusively at  
CC high levels in the skin.  
CC -1- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ALOPECIA UNIVERSALIS  
CC (ALUNC), A RARE AUTOSOMAL RECESSIVE FORM OF HAIR LOSS. IT IS  
CC CHARACTERIZED BY HAIR FOLLICLES WITHOUT HAIR.  
CC -1- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ATRICHA WITH PAPULAR  
CC LESIONS (APL) (ALSO KNOWN AS CONGENITAL ATRICHA). THIS AUTOSOMAL  
CC RECESSIVE DISEASE IS CHARACTERIZED BY PAPULAR LESIONS OVER MOST  
CC OF THE BODY AND ALMOST COMPLETE ABSENCE OF HAIR.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF039196; AAC32258.2; -;  
DR EMBL: AJ277249; CAB87577.2; -;  
DR EMBL: AJ277250; CAB87577.2; JOINED.  
DR EMBL: AJ277251; CAB87577.2; JOINED.  
DR EMBL: AJ277252; CAB87577.2; JOINED.  
DR EMBL: AJ277253; CAB87577.2; JOINED.  
DR EMBL: AJ400825; CAB87577.2; JOINED.  
DR EMBL: AJ400826; CAB87577.2; JOINED.  
DR EMBL: AJ400827; CAB87577.2; JOINED.  
DR EMBL: AJ400828; CAB87577.2; JOINED.  
DR EMBL: AJ400829; CAB87577.2; JOINED.  
DR EMBL: AJ400830; CAB87577.2; JOINED.  
DR EMBL: AJ400831; CAB87577.2; JOINED.  
DR EMBL: AJ400832; CAB87577.2; JOINED.  
DR EMBL: AJ400833; CAB87577.2; JOINED.  
DR EMBL: AJ400834; CAB87577.2; JOINED.  
DR EMBL: AJ400835; CAB87577.2; JOINED.  
DR EMBL: AJ400836; CAB87577.2; JOINED.  
DR EMBL: AJ400837; CAB87577.2; JOINED.  
DR EMBL: AJ277165; CAB86602.1; -;  
DR MIM: 602302; -;  
DR MIM: 203655; -;  
DR MIM: 209500; -;  
DR InterPro: IPR003347; JmjC.  
DR Pfam: PF02373; JmjC; 1.  
KW Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;  
KW Metal-binding; Alternative splicing; Disease mutation.  
FT ZN-FING 600 625  
FT VARSPLIC 1072 1126  
FT VARIANT 620 1126  
FT VARIANT 1022 1022  
FT VARIANT 1136 1136  
FT CONFLICT 337 337 D->G (IN REF. 3).  
FT CONFLICT 446 446 L->W (IN REF. 3).  
FT CONFLICT 584 584 S->G (IN REF. 3).  
SQ SEQUENCE 1189 AA; 127509 MW; 5E244858716B5DF CRC64;  
OY 3 VPWTAW 8  
DB 274 VPWTAW 279  
RESULT 3  
C71N\_ARATH STANDARD: PRT: 483 AA.  
AC Q9STL0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cytochrome P450 71A23 (EC 1.14.-.-).  
GN CYP71A23 OR AT3G48340 OR T29H1.180.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;  
OC Ericaceae; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA:
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unsel'd M.,
RA Farman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Deleney M., Boutry M., Griwell L.A., Mache R., Pulgdenesch P.,
RA De Simone V., Choisme N., Artiguenave F., Robert C., Brotlier P.,
RA Winkler P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T., Nordstiek G.,
RA Relchelt J., Schaefer M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,
RA Cooke R., Landie M., Berger-Liauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Montfort A., Argitrou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shimp S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana."
RA Nature 408:820-822(2000).
RT -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL049659; CAB41168.1; -
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 1 21
FT BINDING 429 429 HEME (BY SIMILARITY).
FT SEQUENCE 483 AA; 54761 MW; 4CB55E1F4AC436C7 CRC64;
SQ

```

```

Query Match 74.5%; Score 38; DB 1; Length 483;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 SRVPTW 8
DB 211 SYVPTW 218

```

```

RESULT 4
C710.ARATH STANDARD; PRT; 489 AA.
AC 09STR7;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DE Cytochrome P450 71A26 (EC 1.14.-.-).
CN CYP71A26 OR AT3G48270 OR T29H1.210.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA:
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unsel'd M.,
RA Farman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Deleney M., Boutry M., Griwell L.A., Mache R., Pulgdenesch P.,
RA De Simone V., Choisme N., Artiguenave F., Robert C., Brotlier P.,
RA Winkler P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T., Nordstiek G.,
RA Relchelt J., Schaefer M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,
RA Cooke R., Landie M., Berger-Liauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Montfort A., Argitrou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shimp S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana."
RA Nature 408:820-822(2000).
RT -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL049659; CAB41171.1; -
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 1 21
FT BINDING 431 431 HEME (BY SIMILARITY).
FT SEQUENCE 489 AA; 55832 MW; 224FC596FB3BA75 CRC64;
SQ

```

```

Query Match 74.5%; Score 38; DB 1; Length 489;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 SRVPTW 8
DB 213 SYVPTW 220

```

```

RESULT 5
REFX5.HUMAN STANDARD; PRT; 616 AA.
AC P48382;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE DNA-binding protein RFX5 (Regulatory factor X subunit 5).  
 GN RFX5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RA MEDLINE=95262896; PubMed=7744245;  
 RA Steimle V., Durand B., Barras E., Zufferey M., Hadam M.R.,  
 RA Mach B., Reich M.;  
 RT "A novel DNA-binding regulatory factor is mutated in primary MHC  
 RL class II deficiency (bare lymphocyte syndrome).";  
 RL Genes Dev. 9:1021-1032(1995).  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Lymphoblast;  
 RA MEDLINE=91170284; PubMed=10072068;  
 RA Nagatajan U.M., Louis-Pence P., Desandro A., Nilsen R., Bushey A.,  
 RA Boss J.M.;  
 RT "RFX-B is the gene responsible for the most common cause of the bare  
 RL lymphocyte syndrome, an MHC class II immunodeficiency.";  
 RL Immunity 10:153-162(1999).  
 RN [3]  
 RP ERRATUM.  
 RA Nagatajan U.M., Louis-Pence P., Desandro A., Nilsen R., Bushey A.,  
 RA Boss J.M.;  
 RL Immunity 10:399-399(1999).  
 RN [4]  
 RP CHARACTERIZATION.  
 RA MEDLINE=20242030; PubMed=10779326;  
 RA Villard J., Peretti M., Masternak K., Barras E., Caretti G.,  
 RA Mantovani R., Reich M.;  
 RT "A functionally essential domain of RFX5 mediates activation of major  
 RL histocompatibility complex class II promoters by promoting  
 RT cooperative binding between RFX and NF-Y.";  
 RL Mol. Cell Biol. 20:3364-3376(2000)  
 CC -1- FUNCTION: ACTIVATES TRANSCRIPTION FROM CLASS II MHC PROMOTERS.  
 CC RECOGNIZES X-BOXES. MEDIATES COOPERATIVE BINDING BETWEEN RFX AND  
 CC NF-Y. RFX BINDS THE X1 BOX OF MHC-II PROMOTERS.  
 CC -1- SUBUNIT: RFX CONSISTS OF AT LEAST THREE DIFFERENT SUBUNITS. RFXAP,  
 CC RFX5 AND RFX-B/REXANK; WITH EACH SUBUNIT REPRESENTING A SEPARATE  
 CC COMPLEMENTATION GROUP. RFX FORMS COOPERATIVE DNA BINDING COMPLEXES  
 CC WITH X2B2 AND CBF/NF-Y. RFX ASSOCIATES WITH CLITA TO FORM AN  
 CC ACTIVE TRANSCRIPTIONAL COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: Nucleus.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
 CC -1- DOMAIN: THE N-TERMINUS IS REQUIRED FOR ITS ASSOCIATION WITH RFXANK  
 CC AND RFXAP, FOR ASSEMBLY OF THE RFX COMPLEX, AND FOR BINDING OF  
 CC THIS COMPLEX TO ITS X BOX TARGET SITE IN THE MHC-II PROMOTER. THE  
 CC C-TERMINUS MEDIATES COOPERATIVE BINDING BETWEEN THE RFX COMPLEX  
 CC AND NF-Y.  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -1- DISEASE: DEFECTS IN RFX5 ARE A CAUSE OF HEREDITARY MHC CLASS II  
 CC DEFICIENCY (ALSO KNOWN AS BARE LYMPHOCYTE SYNDROME (BLS) OR HLA  
 CC CLASS II DEFICIENT COMBINED IMMUNODEFICIENCY); A FORM OF SEVERE  
 CC COMBINED IMMUNODEFICIENCY DISEASE (SCID). RFX5 IS LINKED WITH  
 CC BLS COMPLEMENTATION GROUPS C.  
 CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X85786; CA59771.1; -.  
 DR TRANSFAC; T01672; -.  
 DR MIM; 601863; -.  
 DR MIM; 209920; -.

DR InterPro: IPR003150. RFX-DNA-binding.  
 DR Pfam; PF02257; RFX-DNA-binding; 1.  
 KW DNA-binding; Transcription regulation; Activator; Nuclear protein;  
 KW SCID; Phosphorylation.  
 FT DNA\_BIND 92 168  
 SO SEQUENCE 616 AA; 65322 MW; 5EBB33C677BB717F CRC64;  
 Query Match 74.5%; Score 38; DB 1; Length 616;  
 Best Local Similarity 62.5%; Pred. No. 37;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 SRVPTW 8  
 Db 496 SRLPWTW 503  
 RESULT 6  
 FSPO\_XENLA STANDARD; PRT; 803 AA.  
 ID FSPO\_XENLA  
 AC P35447;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-1999 (Rel. 38, Last annotation update)  
 DE F-spodin precursor.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 CC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93376785; PubMed=8367492;  
 RA Ruiz i Altaba A., Cox C., Jessell T.M., Kiar A.;  
 RT "Ectopic neural expression of a floor plate marker in frog embryos  
 RL injected with the midline transcription factor Pintallavis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).  
 CC -1- FUNCTION: PROMOTES THE ATTACHMENT OF SPINAL CORD AND SENSORY  
 CC NEURON CELLS AND THE OUTGROWTH OF NEURITES IN VITRO. MAY  
 CC CONTRIBUTE TO THE GROWTH AND GUIDANCE OF AXONS IN BOTH THE SPINAL  
 CC CORD AND THE PNS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE FLOOR PLATE.  
 CC -1- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L09123; AA19105.1; -.  
 DR PIR; A47723; A47723.  
 DR InterPro; IPR002861; Reeler.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF02014; Reeler; 1.  
 DR Pfam; PF00090; tsp.1; 6.  
 DR SMART; SM00209; TSP1; 6.  
 DR PROSITE; PSS0092; TSP1; 6.  
 KW Glycoprotein; Signal; Repeat; Cell adhesion.  
 FT SIGNAL 1 23  
 FT CHAIN 24 803  
 FT DOMAIN 437 488 TSP TYPE-1.1.  
 FT DOMAIN 496 548 TSP TYPE-1.2.  
 FT DOMAIN 553 604 TSP TYPE-1.3.  
 FT DOMAIN 609 661 TSP TYPE-1.4.  
 FT DOMAIN 665 716 TSP TYPE-1.5.  
 FT DOMAIN 751 803 TSP TYPE-1.6.  
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 803 AA: 90702 MW: D3A54E329548AED9 CRC64;

Query Match 74.5%; Score 38; DB 1; Length 803;

Best Local Similarity 100.0%; Pred. No. 48;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
DB 755 PWTAW 759

```

RESULT 7
FSPO_RAT STANDARD: PRT: 807 AA.
AC P35446:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE F-spondin precursor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic floor plate;
RX MEDLINE=92208952; PubMed=1555244;
RA Klar A., Baldassare M., Jessell T.M.;
RT "F-spondin: a gene expressed at high levels in the floor plate
RT encodes a secreted protein that promotes neural cell adhesion and
RT neurite extension."
RL Cell 69:95-110(1992).
CC -1- FUNCTION: PROMOTES THE ATTACHMENT OF SPINAL CORD AND SENSORY
CC NEURON CELLS AND THE OUTGROWTH OF NEURITES IN VITRO. MAY
CC CONTRIBUTE TO THE GROWTH AND GUIDANCE OF AXONS IN BOTH THE SPINAL
CC CORD AND THE PNS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE FLOOR PLATE.
CC -1- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M88469; AAA41174.1; -.
DR PIR: A38152; A38152.
DR InterPro: IPR002861; Reeler.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF02014; Reeler1.1.
DR Pfam: PF00090; tsp-1; 6.
DR SMART: SM00209; TSP1; 6.
DR PROSITE: PS50092; TSP1; 5.
KM Glycoprotein; Signal; Repeat; Cell adhesion.
FT SIGNAL 1 28
FT CHAIN 29 807
FT DOMAIN 443 494 TSP TYPE-1 1.
FT DOMAIN 502 554 TSP TYPE-1 2.
FT DOMAIN 559 610 TSP TYPE-1 3.
FT DOMAIN 615 665 TSP TYPE-1 4.
FT DOMAIN 669 720 TSP TYPE-1 5.
FT DOMAIN 755 807 TSP TYPE-1 6.
FT CARBOHYD 214 681 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 681 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 807 AA: 90773 MW: 309525F9EAFEAB9A CRC64;

```

Query Match 74.5%; Score 38; DB 1; Length 807;  
Best Local Similarity 100.0%; Pred. No. 48;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
DB 759 PWTAW 763

```

RESULT 8
AMPN_ECOLI STANDARD: PRT: 869 AA.
ID AMPN_ECOLI
AC P04825;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aminopeptidase N (EC 3.4.11.2) (Alpha-aminocysteine hydrolase).
GN PEPN OR B0932.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87192020; PubMed=2436977;
RA Fogliano M., Garbri S., Lazdunski A.;
RT "Nucleotide sequence of the pepN gene encoding aminopeptidase N of
RT Escherichia coli."
RL Gene 49:303-309(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=87163509; PubMed=3549459;
RA McCaman M.T., Gabe J.D.;
RT "The nucleotide sequence of the pepN gene and its over-expression in
RT Escherichia coli."
RL Gene 48:145-153(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 1-241 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=86310300; PubMed=3018440;
RA McCaman M.T., Gabe J.D.;
RT "Sequence of the promoter and 5' coding region of pepN, and the
RT amino-terminus of peptidase N from Escherichia coli K-12."
WL Gen. Genet. 204:148-152(1986).
RN [6]
RP SEQUENCE OF 1-176 FROM N.A., AND SEQUENCE OF 1-21.
RX MEDLINE=86164315; PubMed=2869947;
RA Bally M., Fogliano M., Bruschi M., Murgier M., Lazdunski A.;
RT "Nucleotide sequence of the promoter and amino-terminal encoding
RT region of the Escherichia coli pepN gene."
RL Eur. J. Biochem. 155:565-569(1986).

```

```

CC -1- FUNCTION: AMINOPEPTIDASE N IS INVOLVED IN THE DEGRADATION OF
CC INTRACELLULAR PEPTIDES GENERATED BY PROTEIN BREAKDOWN DURING
CC NORMAL GROWTH AS WELL AS IN RESPONSE TO NUTRIENT STARVATION.
CC -1- COFACTOR: BINDS 1 ZINC ION.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, BOUND TO THE INNER FACE OF
CC THE CYTOPLASMIC MEMBRANE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1.
CC -----
CC CC SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04020; CAA27647.1; -
DR EMBL: X03709; CAA27336.1; -
DR EMBL: M15676; AAA24318.1; -
DR EMBL: A8000195; AAC74018.1; -
DR EMBL: D90731; BAA35684.1; -
DR EMBL: D90732; BAA35687.1; -
DR EMBL: M15273; AAA24317.1; -
DR PIR: A29045; DPECN.
DR PIR: A27164; A27164.
DR MEROPS: M01.005; -.
DR Ecogene; EG10696; pepN.
DR InterPro; IPR001930; Zn_MTPeptase.
DR InterPro; IPR000130; Zn_MTPeptase.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metallopeptase; Amino-peptidase; Zinc; Membrane;
KW Complete proteome.
FT METAL 296 296 ZINC (CATALYTIC) (BY SIMILARITY).
FT INT_MET 0 0
FT ACT_SITE 297 297 BY SIMILARITY.
FT METAL 300 300 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 319 319 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 380 380 PROTON DONOR (POTENTIAL).
FT CONFLICT 75 75 E -> D (IN REF. 6).
SQ SEQUENCE 869 AA; 98787 MW; 9DED2712669CECA3 CRC64;

Query Match 74.5%; Score 38; DB 1; Length 869;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PWTAW 8
Db 69 PWTAW 73

RESULT 9
C71P_ARATH STANDARD; PRT; 490 AA.
ID C71P_ARATH
AC Q9STK8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Cytochrome P450 71A25 (EC 1.14.-.-).
GN CYP71A25 OR AT3G48360 OR T29H1.200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; Pubmed=11130713;
RA Sahnounbat M., Lemcke K., Rieger M., Ansoorge W., Unseld M.,
Fatmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,

```

```

RA Delseny M., Boutry M., Grivell L.A., Mache R., Pindomenech P.,
RA De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Queller F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Nordstiel B.,
RA Conrad A., Hornischer K., Kauer G., Loehmet T.-H., Nordstiel G.,
RA Reichelt J., Scharte M., Schoen O., Bargues M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Lande A., Berger-Llauró C., Punelle B., Masny D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Cascuberta E.,
RA Montfort A., Argirrou A., Flores M., Liquti R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rüd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Ullrich T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nieman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Nakayama A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL049659; CAB41170.1; -
DR InterPro; IPR001128; Cyt_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 1 21 POTENTIAL.
FT BINDING 431 431 HEME (BY SIMILARITY).
SQ SEQUENCE 490 AA; 55783 MW; FB97BA80A4C1EA50 CRC64;

Query Match 72.5%; Score 37; DB 1; Length 490;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SRVPTAW 8
Db 212 SRVPTAW 219

RESULT 10
POLG_EMCV STANDARD; PRT; 2290 AA.
ID POLG_EMCV
AC P03304;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C, P3A; Genome-linked protein VPg; Picornain 3C
DE (EC 3.4.22.28) (Protease 3C) (P3c); RNA-directed RNA polymerase P2D
DE (EC 2.7.7.48)].
OS Encephalomyocarditis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Cardiovirus.
OX NCBI_TaxID=12104;
RN [1]

```



```

RP SEQUENCE FROM N.A.
RX MEDLINE=84169586; Pubmed=6324136;
RA Palmenberg A.C., Kirby E.M., Janda M.R., Drake N.L., Duke G.M.,
RA Portratz K.F., Collett M.S.;
RT "The nucleotide and deduced amino acid sequences of the
RT encephalomyocarditis viral polypeptide coding region.";
RL Nucleic Acids Res. 12:2969-2985(1984).
CC -1- FUNCTION: PC POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00463; CAA25152.1; -.
DR PIR; A03906; GNNE.
DR HSSP; P12296; 2MEV.
DR MEROPS; C03.009; -.
DR MEROPS; U29.001; -.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR001676; Riv.
DR Pfam; PF00073; Ihv; 3.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein: Coat protein: Core protein: Transferase:
KW RNA-directed RNA polymerase: Hydrolase; Thiol protease; Myristate.
FT PROPEP 1 67 LEADER PEPTIDE.
FT CHAIN 68 136 COAT PROTEIN VP4 (RHO).
FT CHAIN 137 391 COAT PROTEIN VP2 (BETA).
FT CHAIN 392 622 COAT PROTEIN VP3 (GAMMA).
FT CHAIN 623 910 COAT PROTEIN VP1 (ALPHA).
FT CHAIN 911 1056 CORE PROTEIN P2A (G).
FT CHAIN 1057 1192 CORE PROTEIN P2B (I).
FT CHAIN 1193 1517 CORE PROTEIN P2C (F).
FT CHAIN 1518 1605 CORE PROTEIN P3A.
FT CHAIN 1606 1625 GENOME-LINKED PROTEIN VP6 (H).
FT CHAIN 1626 1830 PICORNAIN 3C (P22).
FT CHAIN 1831 2290 RNA-DIRECTED RNA POLYMERASE P3D (E).
FT LIPID 68 68 MYRISTATE (BY SIMILARITY).
FT ACCT_SITE 1784 1784 PROTEASE (POTENTIAL).
FT ACCT_SITE 1802 1802 PROTEASE (POTENTIAL).
SQ SEQUENCE 2290 AA; 255756 MW; 26BC81BB7CF68CB5 CRC64;

Query Match 72.5%; Score 37; DB 1; Length 2290;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 SRVPTAW 8
DB 964 SRAPNPMW 971

RESULT 11
EXOV_RHIME STANDARD; PRT; 316 AA.
AC P33701.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Succinoglycan biosynthesis protein exov.
GN EXOV OR R81072 OR SMB20349.
OS Rhizobium meliloti (Sinorhizobium meliloti).

```

```

OG Plasmid pSymb (megaplasmid 2).
CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Sinorhizobium.
OX NCB1_TaxID=382;
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RC MEDLINE=94042869; Pubmed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT succinoglycan by Rhizobium meliloti.";
RL J. Bacteriol. 175:7033-7044(1993).
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RC MEDLINE=94042870; Pubmed=8226646;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Genes needed for the modification, polymerization, export, and
RT processing of succinoglycan by Rhizobium meliloti: a model for
RT succinoglycan biosynthesis.";
RL J. Bacteriol. 175:7045-7055(1993).
RP SEQUENCE FROM N.A.
RX STRAIN=RCR2011 / S047;
RC MEDLINE=94162682; Pubmed=8118055;
RA Becker A., Kleckmann A., Kuester H., Keller M., Arnold W.,
RA Puhler A.;
RT "Analysis of the Rhizobium meliloti genes exov, exow, exot and
RT exoi involved in exopolysaccharide biosynthesis and nodule invasion:
RT exoi and exow probably encode glucosyltransferases.";
RL Mol. Plant Microbe Interact. 6:735-744(1993).
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RC MEDLINE=21396508; Pubmed=11481431;
RA Flann T.M., Weidner S., Wong K., Bunmester J., Chain P.,
RA Vorhölter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
RP -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20758; AAA16052.1; ALT_INT.
DR EMBL; Z22646; CAA80360.1; -.
DR EMBL; Z22646; CAA80360.1; -.
DR PIR; A49349; A49349.
KW Exopolysaccharide synthesis; Plasmid; Complete proteome.
SQ SEQUENCE 316 AA; 35321 MW; 9BA45C6E7D763171 CRC64;

Query Match 70.6%; Score 36; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 RVPWTA 7
DB 181 RVPWTA 186

RESULT 12
AMFR_HUMAN STANDARD; PRT; 323 AA.
AC P26442.
DT 01-AUG-1992 (Rel. 23, Created)

```

```

DT 01-AUG-1992 (Rel. 23, last sequence update)
DE 16-OCT-2001 (Rel. 40, last annotation update)
DE Autocrine motility factor receptor precursor (AMF receptor) (GP78).
GN AMF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91302382; PubMed=1649192;
RA Matarabe H., Carmi P., Hogan V., Raz T., Silletti S., Nabi I.R.,
RA Raz A.;
RT "Purification of human tumor cell autocrine motility factor and
RT molecular cloning of its receptor.";
RL J. Biol. Chem. 266:13442-13448(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta; PubMed=7626106;
RA Huang B., Xie Y., Raz A.;
RT "Identification of an upstream region that controls the transcription
RT of the human autocrine motility factor receptor.";
RL Biochem. Biophys. Res. Commun. 212:727-742(1995).
CC -1- FUNCTION: SPECIFIC RECEPTOR FOR THE AUTOCRINE MOTILITY FACTOR.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: PHOSPHORYLATED IN THE PRESENCE OF AMF.
CC -1- PTM: O-GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63175; AAA3671.1; -.
DR EMBL; L35233; AAA9362.1; -.
DR PIR; A39877; A39877.
DR MIM; 603243; -.
KW Receptor; Glycoprotein; Phosphorylation; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 323
FT DOMAIN 18 110 AUTOCRINE MOTILITY FACTOR RECEPTOR.
FT TRANSMEM 111 137 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 138 323 POTENTIAL.
FT CARBOHYD 24 24 CYTOPLASMIC (POTENTIAL).
FT MOD. RES 194 194 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 323 AA; 34325 MW; 0A7AF4DCF90A8700 CRC64;

```

```

Query Match          70.6%; Score 36; DB 1; Length 323;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 SRVPTWAM 8
Db 149 SGVDMTAM 156

```

```

RESULT 13
YXN1_YEAST
ID YXN1_YEAST STANDARD; PRT; 351 AA.
AC P53860;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 40.7 kDa protein in CSL4-ORE2 intergenic region.
GN YNL231C OR N1158.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

```

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051596; PubMed=8896273;
RA Pandolfo D., de Antoni A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
RT reading frames including a novel gene encoding a globin-like
RT domain.";
RL Yeast 12:1071-1076(1996).
CC -1- SIMILARITY: BELONGS TO THE SEC14 CYTOSOLIC FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z69381; CAA9367.1; -.
DR EMBL; Z71507; CAA96136.1; -.
DR SGD; S0005175; YNL231C.
DR InterPro; IPR001251; CRAL_TRIO.
DR Pfam; PF00650; CRAL_TRIO; 1.
DR SMART; SM00516; SEC14; 1.
KW Hypothetical protein.
SQ SEQUENCE 351 AA; 40714 MW; 24C5B3262016F037 CRC64;

```

```

Query Match          70.6%; Score 36; DB 1; Length 351;
Best Local Similarity 50.0%; Pred. No. 47;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 SRVPTWAM 8
Db 243 TNPFWLAM 250

```

```

RESULT 14
SYT8_MOUSE
ID SYT8_MOUSE STANDARD; PRT; 395 AA.
AC O9R0N6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Synaptotagmin VIII (SytVIII).
GN SYT8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX SRRAIN=ICR; TISSUE=Cerebellum;
RX MEDLINE=20002669; PubMed=10531343;
RA Fukuda M., Kanno E., Mikoshiba K.;
RT "Conserved N-terminal cysteine motif is essential for homo- and
RT heterodimer formation of synaptotagmins III, V, VI, and X.";
RL J. Biol. Chem. 274:31421-31427(1999).
CC -1- FUNCTION: MAY BE INVOLVED IN CA2+-DEPENDENT EXOCYTOSIS OF
CC SECRETORY VESICLES THROUGH CA2+ AND PHOSPHOLIPID BINDING TO THE C2
CC DOMAIN OR MAY SERVE AS CA2+ SENSORS IN THE PROCESS OF VESICULAR
CC TRAFFICKING AND EXOCYTOSIS (BY SIMILARITY).
CC -1- SUBUNIT: Homodimer. Can also forms heterodimer (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; SYNAPTIC
CC VESICLES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SYNAPTOTAGMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL: AB026805; BAAB5777.1; -.  
 DR HSSP: P21707; IRSY.  
 DR MGD: MGI:1859867; SYL8.  
 DR InterPro: IPR000008; C2.  
 DR InterPro: IPR002149; LRI.  
 DR InterPro: IPR001565; Synaptotagmin.  
 DR Pfam: PF00168; C2; 2.  
 DR PRINTS: PR00399; SYNAPTOTAGMN.  
 DR SMART: SM00239; C2; 2.  
 DR PROSITE: PS00499; C2\_DOMAIN\_1; FALSE\_NEG.  
 DR PROSITE: PS50004; C2\_DOMAIN\_2; 2.  
 KW Transmembrane; Repeat; Synapse.  
 FT DOMAIN 1 44 VESICULAR (POTENTIAL).  
 FT TRANSMEM 45 65 POTENTIAL.  
 FT DOMAIN 66 395 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 115 213 C2\_DOMAIN\_1.  
 FT DOMAIN 243 342 C2\_DOMAIN\_2.  
 SQ SEQUENCE 395 AA; 44093 MW; 10E833FA8C454A3B CRC64;

Query Match 70.6%; Score 36; DB 1; Length 395;  
 Best Local Similarity 50.0%; Pred. No. 52;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRVPTAW 8  
 : : : : :  
 Db 36 TRIPWRW 43

RESULT 15  
 C71M\_ARATH STANDARD; PRT; 490 AA.  
 AC Q9STL1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome P450 71A22 (EC 1.14.-.-).  
 GN CYP71A22 OR AT3G48330 OR T29H11.170.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RX STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016720; PubMed=1130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansgorge W., Unseld M.,  
 RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Deiseny M., Boutry M., Griwell L.A., Macho R., Puigdomenech P.,  
 RA De Simone V., Choisme N., Arltgenave F., Robert C., Brottier P.,  
 RA Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Mumbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,  
 RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordliek G.,  
 RA Reichelt J., Scharte M., Schoen O., Barges M., Terol J., Climent J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,  
 RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,  
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 RA Monfort A., Argitlou A., Flores M., Liguori R., Vitale D.,  
 RA Mannhaupt G., Haese D., Schoof H., Rood H.L., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Maltis A., Uterback T., Fujii C.Y., Shea T.P.,  
 RA Creasy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Militscher J., Sellers P., Gill J.E., Feldlyum T.V.,  
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,

RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakayama A., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Tanabe A., Yamada M., Yasuda M., Yabata S.,  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana."  
 RL Nature 408:820-822(2000).  
 CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL049659; CAB41167.1; -.  
 DR InterPro: IPR001128; CYL\_P450.  
 DR Pfam: PF00067; P450; 1.  
 DR PRINTS: PR00385; P450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.  
 FT TRANSMEM 2 22 POTENTIAL.  
 FT BINDING 432 432 HEME (BY SIMILARITY).  
 SQ SEQUENCE 490 AA; 55998 MW; 64C96AF349CA5672 CRC64;

Query Match 70.6%; Score 36; DB 1; Length 490;  
 Best Local Similarity 83.3%; Pred. No. 63;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 VPTAW 8  
 : : : : :  
 Db 216 VPTAW 221

Search completed: August 15, 2002, 11:58:16  
 Job time: 516 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:49:05 ; Search time 58.57 Seconds  
(without alignments)  
23.629 Million cell updates/sec

Title: US-09-613-092a-5\_COPY\_3\_10  
Perfect score: 51  
Sequence: 1 SRVPM7AW 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	87.4	81	5 Q9VBC3	Q9vbc3 drosophila
2	41	80.4	164	10 Q9FS18	Q9fs18 hordeum vul
3	39	76.5	335	16 Q9WX57	Q9wx57 thermotoga
4	39	76.5	668	2 Q9ALX8	Q9alx8 burkholderi
5	39	76.5	797	16 Q98P32	Q98p32 rhizobium 1
6	39	76.5	800	4 Q96H33	Q96h33 homo sapien
7	38	74.5	147	11 Q614Z7	Q614z7 mus musculu
8	38	74.5	151	11 Q9DA47	Q9da47 mus musculu
9	38	74.5	162	10 Q9M042	Q9m042 arabidopsis
10	38	74.5	271	11 Q9CTW8	Q9ctw8 mus musculu
11	38	74.5	297	10 Q9FTE7	Q9fte7 oryza sativ
12	38	74.5	434	4 Q9UG77	Q9ug77 homo sapien
13	38	74.5	441	16 Q9HYE2	Q9hye2 pseudomonas
14	38	74.5	454	10 Q9X159	Q9x159 arabidopsis
15	38	74.5	595	11 Q91VH2	Q91vh2 mus musculu
16	38	74.5	624	4 Q94862	Q94862 homo sapien

17	38	74.5	773	10 Q9SHP2	Q9shp2 arabidopsis
18	38	74.5	802	13 Q9W770	Q9w770 gallus gall
19	38	74.5	807	4 Q9HC66	Q9hc66 homo sapien
20	38	74.5	807	6 Q9GLX9	Q9glx9 bos taurus
21	37	72.5	309	10 Q9SSR0	Q9ssr0 arabidopsis
22	37	72.5	415	2 Q937Z8	Q937z8 mycobacteri
23	37	72.5	602	3 Q9P8C3	Q9p8c3 acromonium
24	37	72.5	2292	12 Q66765	Q66765 encephalomy
25	37	72.5	2292	12 Q66850	Q66850 encephalomy
26	36	70.6	97	4 Q9H2T5	Q9h2t5 homo sapien
27	36	70.6	241	13 Q9DDW2	Q9ddw2 tuiga rudrip
28	36	70.6	247	16 Q92R85	Q92r85 rhizobium m
29	36	70.6	291	16 P74129	P74129 synechocyst
30	36	70.6	328	3 Q9UR34	Q9ur34 schizosacch
31	36	70.6	335	6 Q97916	Q97916 bos taurus
32	36	70.6	437	10 Q9M1D6	Q9m1d6 arabidopsis
33	36	70.6	530	4 Q13161	Q13161 homo sapien
34	36	70.6	541	16 Q9A4N7	Q9a4n7 caulobacter
35	36	70.6	657	2 Q59446	Q59446 fibrobacter
36	36	70.6	669	2 Q59445	Q59445 fibrobacter
37	36	70.6	823	5 Q25343	Q25343 leishmania
38	36	70.6	851	10 Q9M1F2	Q9m1f2 arabidopsis
39	36	70.6	870	10 Q96376	Q96376 clarkia bre
40	36	70.6	871	10 Q9ZPM5	Q9zpm5 clarkia con
41	36	70.6	1524	10 Q9S1Q5	Q9s1q5 arabidopsis
42	36	70.6	1750	10 Q9SKJ4	Q9skj4 arabidopsis
43	35	68.6	109	4 Q9GZM3	Q9gzm3 homo sapien
44	35	68.6	182	5 Q22231	Q22231 caenothadl
45	35	68.6	254	10 Q9LTR7	Q9ltr7 arabidopsis

## ALIGNMENTS

RESULT	ID	Q9VBC3	PRELIMINARY:	PRT:	81 AA.
1	AC	Q9VBC3	01-MAY-2000 (TREMblrel. 13, Created)		
	DT	01-MAY-2000 (TREMblrel. 13, Last sequence update)			
	DT	01-DEC-2001 (TREMblrel. 19, Last annotation update)			
	DE	CG14244 PROTEIN.			
	GN	CG14244.			
	OS	Drosophila melanogaster (Fruit fly).			
	OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
	OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
	OX	NCBI_TaxID=7227;			
	RM	[1]			
	RP	SEQUENCE FROM N.A.			
	RC	STRAIN-BERKELEY:			
	RX	MEDLINE-20196006; PubMed-10731132;			
	RA	Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
	RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
	RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
	RA	Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
	RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe C.R., Pfiffer B.D.,			
	RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikiotis G.L.G.,			
	RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
	RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
	RA	Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,			
	RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,			
	RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
	RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
	RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
	RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
	RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,			
	RA	Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,			
	RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
	RA	Harlid N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
	RA	Hoslin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,			
	RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
	RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkalov G., Mishina N.V., Mobarly C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Stector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarum D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003757; AAF56619.1; -;  
 DR FlyBase: FBgn0040607; CG14244.  
 DR InterPro: IPR002557; Chitin\_binding.  
 DR Pfam: PF01607; Chitin\_bind\_2; 1.  
 SQ SEQUENCE 81 AA; 9185 MW; AF1E54DECEC8CE1 CRC64;

Query Match 82.4%; Score 42; DB 5; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPMWTAW 8  
 |||||  
 Db 58 VPMWTAW 63

RESULT 2  
 ID 09FS18 PRELIMINARY; PRT; 164 AA.  
 AC 09FS18;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE COLD-REGULATED PROTEIN.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OC NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LEAF;  
 RA Cattivelli L.;  
 RT "Plasid development and plastoquinone redox state control the low  
 temperature specific accumulation of COR14b protein in barley.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ291295; CAC12881.1; -;  
 DR HSSP: P12994; 1FTJ.  
 SQ SEQUENCE 164 AA; 17613 MW; 1F5A99FF0D59B746 CRC64;

Query Match 80.4%; Score 41; DB 10; Length 164;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RVPMTAW 8  
 |||||  
 Db 63 RVPMTAW 69

RESULT 3  
 ID 09WX57 PRELIMINARY; PRT; 335 AA.  
 AC 09WX57;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE OLIGOPEPTIDE ABC TRANSPORTER, PERMEASE PROTEIN.  
 GN TM0072.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OC NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB / DSM 13109;  
 RX MEDLINE=99287316; Pubmed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 genome sequence of *Thermotoga maritima*.";  
 RL Nature 391:323-329(1999).  
 DR EMBL: AE001694; AAD35166.1; -;  
 DR TIGR: TM0072;  
 DR InterPro: IPR000515; BPD\_transp.  
 DR Pfam: PF00528; BPD\_transp; 1.  
 KN Complete proteome.  
 SQ SEQUENCE 335 AA; 38222 MW; 653D1AA484DC9FF3 CRC64;

Query Match 76.5%; Score 39; DB 16; Length 335;  
 Best Local Similarity 83.3%; Pred. No. 68;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPMWTAW 8  
 :|||  
 Db 109 LPMWTAW 114

RESULT 4  
 ID 09ALX8 PRELIMINARY; PRT; 668 AA.  
 AC 09ALX8;  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE HYDROGENASE-4 COMPONENT B.  
 OS Burkholderia pseudomallei (Pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC Burkholderia.  
 OC NCBI\_TaxID=28450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=G9313;  
 RA Steiner B., Meyer R., Bowen M., Morrill W.;  
 RT "Random sequencing of Burkholderia pseudomallei strain G9313 for  
 RT clinical PCR development.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIDIONE = NAD(+) + UBIDINOL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC EMBL: AF335723; AAK06855.1; -;  
 DR InterPro: IPR003439; ABC\_transport.  
 DR InterPro: IPR003918; NADhub\_oxidrcse4.  
 DR InterPro: IPR003916; NADhub\_oxidrcse5.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR Pfam: PF00361; Oxidored\_q1; 1.  
 DR PRINTS: PRO1434; NADHDGNASE5.  
 DR PRINTS: PRO1437; NTOXDRDASE4.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 KW NAD: Oxidoreductase; Transmembrane; UNKNOWN\_1.  
 SQ SEQUENCE 668 AA; 71822 MW; 2F016ACB950DD519 CRC64;

Query Match 76.5%; Score 39; DB 2; Length 668;  
 Best Local Similarity 83.3%; Pred. No. 13e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPWTAW 8  
:|||||  
DB 382 PWTAW 387

RESULT 5  
O98P32  
ID O98P32 PRELIMINARY; PRT: 797 AA.  
AC O98P32:  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE FORMATE DEHYDROGENASE.  
GN MLL9633.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Plasmid pMLB.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneo T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
DR EMBL: AP003017; BAB54823.1; -  
DR InterPro: IPR001467; Molybdopterin.  
DR Pfam: PF00384; molybdopterin; 2.  
DR Plasmid; Complete proteome.  
KW SEQUENCE 797 AA; 87533 MW; CA69E6373093B0FE CRC64;  
SQ

Query Match 76.5%; Score 39; DB 16; Length 797;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SRVPTAW 8  
:|||||  
DB 599 TKVPMDAW 606

RESULT 6  
O96H33  
ID O96H33 PRELIMINARY; PRT: 800 AA.  
AC O96H33:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE UNKNOWN (PROTEIN FOR IMAGE:3050359) (FRAGMENT).  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=MELANOMA;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: BC008946; AA08946.1; -  
FT NON\_TER 1  
SQ SEQUENCE 800 AA; 85125 MW; 708926A0A67085E6 CRC64;

Query Match 76.5%; Score 39; DB 4; Length 800;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPWTAW 8  
:|||||  
DB 80 VPWTAW 85

RESULT 7  
O61427  
ID O61427 PRELIMINARY; PRT: 147 AA.  
AC O61427:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ALPHA 1 TYPE I COLLAGEN (FRAGMENT).  
GN COL1A1 OR COL1A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=LIVER;  
RX MEDLINE=94344105; PubMed=8065328;  
RA Rhodes K., Rippe R.A., Umezawa A., Nehls M., Brenner D.A., Breindl M.;  
RT "DNA methylation represses the murine alpha 1(I) collagen promoter by  
RT an indirect mechanism";  
RL Mol. Cell. Biol. 14:5950-5960(1994).  
DR EMBL: X54876; CA938657.1; -  
DR MGD; MGI:88467; Col1a1.  
DR InterPro: IPR001007; VWFc.  
DR Pfam: PF00093; vwc; 1.  
DR SMART: SM00214; VWC; 1.  
DR PROSITE: PS01208; VWFc; 1.  
DR Collagen.  
FT NON\_TER 147  
SQ SEQUENCE 147 AA; 16652 MW; 9263BP0A91B4307D CRC64;

Query Match 74.5%; Score 38; DB 11; Length 147;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
:|||||  
DB 125 PWTAW 129

RESULT 8  
O9DA47  
ID O9DA47 PRELIMINARY; PRT: 151 AA.  
AC O9DA47:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 4933425020RIK PROTEIN.  
GN 4933425020RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schiml L.W., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK006188: BAB24448.1; -  
 DR MGD: MGI:1914016: 4933425020Rik.  
 SQ SEQUENCE 151 AA; 17201 MW; 4B0658DD62565EFD CRC64;

Query Match 74.5%; Score 38; DB 11; Length 151;  
 Best Local Similarity 83.3%; Pred. No. 46;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 VPWTAW 8  
 |||||  
 Db 21 VPMASAW 26

RESULT 9  
 Q9M042 PRELIMINARY; PRT; 162 AA.  
 ID Q9M042;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 17.8 KDA PROTEIN.  
 GN T1008-10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W., Rudd S.,  
 RA Lemcke K., Mayer K.F.X.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA ED Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL161746: CAB81913.1; -  
 DR HSSP: P77366: 1FUX.  
 KW Hypothetical protein.  
 SQ SEQUENCE 162 AA; 17823 MW; B34C3E450B61AD59 CRC64;

Query Match 74.5%; Score 38; DB 10; Length 162;  
 Best Local Similarity 83.3%; Pred. No. 49;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPWTAW 8  
 |||||  
 Db 67 VPMWTW 72

RESULT 10  
 Q9CTW8 PRELIMINARY; PRT; 271 AA.  
 ID Q9CTW8;  
 AC O9CTW8;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 4930565F05RIK PROTEIN (FRAGMENT).  
 GN 4930565F05RIK.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukuishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Mashima T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK019785: BAB31852.1; -  
 DR MGD: MGI:1924880: 4930565F05Rik.  
 FT NON\_TER 1  
 SQ SEQUENCE 271 AA; 29278 MW; 3654CD2B7224A24 CRC64;

Query Match 74.5%; Score 38; DB 11; Length 271;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PWTAW 8  
 |||||  
 Db 209 PWTAW 213

RESULT 11  
 Q9PTE7 PRELIMINARY; PRT; 297 AA.  
 ID Q9PTE7;  
 AC O9PTE7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE P0698A04.4 PROTEIN (P0494A10.20 PROTEIN).  
 GN P0698A04.4 OR P0494A10.20.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Euphorbiaceae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT clone:P0494A10."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP002541: BAB55480.1; -  
 DR InterPro: IPR004277; PSS.



DR Pfam: PF03034; PSS: 1.  
SQ SEQUENCE 297 AA; 34429 MW; E1730EFC71208E09 CRC64;

Query Match 74.5%; Score 38; DB 10; Length 297;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
DB 32 PWTAW 36

RESULT 12  
09UG77

ID 09UG77; PRELIMINARY; PRT; 434 AA.

AC 09UG77; 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE HYPOTHETICAL 45.1 KDA PROTEIN (FRAGMENT).  
GN DKPZP586K091.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=UTERUS;  
RA Ottenwelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL050135; CAB43285.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 434 AA; 45064 MW; FF4BA8CA93693B0 CRC64;

Query Match 74.5%; Score 38; DB 4; Length 434;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SRPWTAW 8  
|||||  
DB 314 SRPWTAW 321

RESULT 13  
09HYE2

ID 09HYE2; PRELIMINARY; PRT; 441 AA.

AC 09HYE2; 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE HYPOTHETICAL PROTEIN PA3464.  
GN PA3464.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Storer C.K., Pham X.-O.T., Ervin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen."  
RL Nature 406:959-964 (2000).  
DR EMBL; AE004767; AAG06852.1; -.

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 441 AA; 49000 MW; 4923F37C2A044DA2 CRC64;

Query Match 74.5%; Score 38; DB 16; Length 441;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
DB 144 PWTAW 148

RESULT 14  
09XI59

ID 09XI59; PRELIMINARY; PRT; 454 AA.

AC 09XI59; 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE F9L1.4 PROTEIN.  
GN F9L1.4.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Lee J., Liu A., Li J., Kremenetskaia I., Lucos J., Gonzalez A.,  
RA Altaji H., Araujo R., Brooks S., Buehler E., Chao Q., Conn L.,  
RA Conway A.B., Dunn P., Hansen N., Huizer L., Khan S., Kim C., Palm C.,  
RA Rowley D., Shinn P., Walker M., Davis R.W., Eckert J.R.,  
RA Federspiel N.A., Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC F9L1 sequence."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC007591; AAD39639.1; -.  
DR InterPro: IPR004277; PSS.  
DR Pfam: PF03034; PSS: 1.  
SQ SEQUENCE 454 AA; 52610 MW; 1EDC8A5A4188A44 CRC64;

Query Match 74.5%; Score 38; DB 10; Length 454;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PWTAW 8  
|||||  
DB 32 PWTAW 36

RESULT 15  
09IVH2

ID 09IVH2; PRELIMINARY; PRT; 595 AA.

AC 09IVH2; 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SIMILAR TO SORTING NEXIN 9.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BREAST TUMOR;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC014814; AAH14814.1; -.  
SQ SEQUENCE 595 AA; 66545 MW; 3D556847662D816D CRC64;

Query Match 74.5%; Score 38; DB 11; Length 595;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 PWTAW 8  
 |||||  
 Db 107 PWTAW 111

Search completed: August 15, 2002, 11:57:51  
 Job time: 526 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:49:02 ; Search time 73.67 Seconds  
(without alignments)  
12.062 Million cell updates/sec

Title: US-09-613-092a-7\_COPY\_3\_10  
Perfect score: 45  
Sequence: 1 RRFVHRRP 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneeq/geneeqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	15	20	AAV30353
2	40	88.9	15	20	AAV30354
3	36	80.0	328	22	AAV41174
4	35	77.8	115	22	ABG25336
5	35	77.8	663	22	ABG64497
6	34.5	76.7	287	20	AAV38511
7	34.5	76.7	304	20	AAV38512
8	34	75.6	70	22	AAV49814
9	34	75.6	156	22	AAV5292
10	34	75.6	156	22	AAV5293
11	34	75.6	217	22	ABV5332

12	34	75.6	243	22	ABV58399	Drosophila melanog
13	34	75.6	273	22	ABG18986	Novel human diagno
14	34	75.6	280	22	ABG10923	Novel human diagno
15	34	75.6	282	22	ABG10922	Novel human diagno
16	34	75.6	303	22	ABG04502	Novel human diagno
17	34	75.6	351	20	AAV59757	Human normal ovari
18	33	73.3	68	22	AAV50169	Proionibacterium
19	33	73.3	95	22	ABG24721	Novel human diagno
20	33	73.3	155	18	AAV10569	Mabinlin II prepro
21	33	73.3	155	18	AAV23587	Mabinlin MBLT fro
22	33	73.3	246	22	AAV42864	Proionibacterium
23	33	73.3	580	20	AAV34665	Chlamydia pneumoni
24	33	73.3	1055	22	AAV39198	Human polypeptide
25	32	71.1	35	21	AAV25737	Human secreted pro
26	32	71.1	52	22	AAV50373	Proionibacterium
27	32	71.1	53	22	AAV57016	Proionibacterium
28	32	71.1	53	22	AAV67050	Proionibacterium
29	32	71.1	65	22	AAV52989	Proionibacterium
30	32	71.1	84	22	AAV52388	Proionibacterium
31	32	71.1	85	21	AAV68825	A. thaliana P1P1 c
32	32	71.1	125	22	AAV02969	Human polypeptide
33	32	71.1	134	21	AAV61646	Arabidopsis thalia
34	32	71.1	141	22	ABG10757	Novel human diagno
35	32	71.1	143	21	AAV61645	Arabidopsis thalia
36	32	71.1	175	21	AAV61644	Arabidopsis thalia
37	32	71.1	184	22	ABG68567	Drosophila melanog
38	32	71.1	193	21	AAV29225	Arabidopsis thalia
39	32	71.1	196	22	AAV95630	Human protein sequ
40	32	71.1	202	21	AAV29224	Arabidopsis thalia
41	32	71.1	202	21	AAV96820	A. thaliana PHO80-
42	32	71.1	222	21	AAV29223	Arabidopsis thalia
43	32	71.1	227	21	AAV25731	Human secreted pro
44	32	71.1	295	21	AAV19761	Arabidopsis thalia
45	32	71.1	295	21	AAV48834	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1	AAV30353	standard; Peptide: 15 AA.
ID	AAV30353	
XX	AAV30353:	
XX	09-NOV-1999 (first entry)	
DE	Epitope derived from pneumococcal surface adhesion A protein.	
XX		
KW	Pneumococcal surface adhesion A protein; Psaa; monoclonal antibody;	
KW	vaccine; Streptococcus pneumoniae Infection.	
XX		
OS	Streptococcus pneumoniae.	
XX		
PN	W09945121-A1.	
XX		
PD	10-SEP-1999.	
XX		
PF	26-FEB-1999; 99WO-US04326.	
XX		
PR	02-MAR-1998; 98US-0076565.	
XX		
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PI	Ades EW, Calzone GM, Sampson JS, Tharpe JA, Westerink MAJ;	
PI	Zeller JL;	
XX		
DR	WPI; 1999-540849/45.	
XX		
PT	New peptides corresponding to Streptococcus pneumoniae Psaa, used	
PT	for treating or preventing Streptococcus pneumoniae infection in a	
PT	subject	
XX		

PS Claim 6; Page 43; 58pp; English.  
 XX  
 CC AAY30351-54 represent immunogenic peptides which are derived from  
 CC a pneumococcal surface adhesion A protein (PsaA). The specification  
 CC describes monoclonal antibodies which bind epitopes of the PsaA protein  
 CC (e.g present sequence). The peptides can be used in vaccines to prevent  
 CC Streptococcus pneumoniae infections. The antibodies of the invention  
 CC can also be used to detect S. pneumoniae in a sample or individual.  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 45; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRP 8  
 |||||  
 Db 3 rrfvhrp 10

RESULT 2  
 AAY30354  
 ID AAY30354 standard; Peptide; 15 AA.  
 XX  
 AC AAY30354;

DT 09-NOV-1999 (first entry)

DE Epitope derived from pneumococcal surface adhesion A protein.

KM Pneumococcal surface adhesion A protein; PsaA; monoclonal antibody;  
 KM vaccine; Streptococcus pneumoniae infection.

OS Streptococcus pneumoniae.

PN WO9945121-A1.

PD 10-SEP-1999.

PF 26-FEB-1999; 99WO-US04326.

PR 02-MAR-1998; 98US-0076565.

(USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerlink MAJ;  
 PI Zeiler JL;

DR WPI; 1999-540849/45.

PT New peptides corresponding to Streptococcus pneumoniae PsaA, used  
 PT for treating or preventing Streptococcus pneumoniae infection in a  
 PT subject

PS Claim 6; Page 43; 58pp; English.

CC AAY30351-54 represent immunogenic peptides which are derived from  
 CC a pneumococcal surface adhesion A protein (PsaA). The specification  
 CC describes monoclonal antibodies which bind epitopes of the PsaA protein  
 CC (e.g present sequence). The peptides can be used in vaccines to prevent  
 CC Streptococcus pneumoniae infections. The antibodies of the invention  
 CC can also be used to detect S. pneumoniae in a sample or individual.  
 XX

SQ Sequence 15 AA;

Query Match 88.9%; Score 40; DB 20; Length 15;  
 Best Local Similarity 87.5%; Pred. No. 0.32;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVHRP 8  
 |||||

Db 3 rrfvhrp 10

RESULT 3  
 AAM41174  
 ID AAM41174 standard; Protein; 328 AA.  
 XX  
 AC AAM41174;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6105.

KM Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KM peripheral nervous system; neuropathy; central nervous system; CNS;  
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KM leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0596042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR WPI; 2001-442253/47.

DR N-PSDB; AAI60330.

PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 6105; 10078pp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX

SQ Sequence 328 AA;

Query Match 80.0%; Score 36; DB 22; Length 328;  
 Best Local Similarity 75.0%; Pred. No. 38;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RRFVRRP 8  
: || ||||  
Db 12 grfchrrp 19

RESULT 4  
ABG25336  
ID ABG25336 standard; Protein: 115 AA.  
XX  
AC ABG25336;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #25327.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSEQ-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS89523.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 55695; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 115 AA;

Query Match 77.8%; Score 35; DB 22; Length 115;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FVHRRP 8  
: |||||  
Db 107 fvhrrp 112

RESULT 5  
ABB64497  
ID ABB64497 standard; Protein: 663 AA.  
XX  
AC ABB64497;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 20283.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL08600.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 20283; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 663 AA;

Query Match 77.8%; Score 35; DB 22; Length 663;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RRFVRRP 8  
: |||||  
Db 369 krfvrrp 376

RESULT 6  
AA38511  
ID AA38511 standard; Protein: 287 AA.  
XX  
AC AA38511;  
XX

```

DT 08-OCT-1999 (first entry)
XX
XX Neisseria gonorrhoeae antigen encoded by partial ORF5.
DE
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
XX Neisseria gonorrhoeae.
OS
XX WO9924578-A2.
XX
XX
XX 20-MAY-1999.
XX
XX
XX 09-OCT-1998; 98WO-IB01665.
XX
XX
XX 01-SEP-1998; 98GB-0019016.
XX 06-NOV-1997; 97GB-0023516.
XX 14-NOV-1997; 97GB-0024190.
XX 18-NOV-1997; 97GB-0024386.
XX 27-NOV-1997; 97GB-0025158.
XX 10-DEC-1997; 97GB-0026147.
XX 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Massignani V, Pizsa M, Rappuoli R, Scarlato V;
XX WPI: 1999-327407/27.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 71; 524pp; English.
XX
XX Amino acid sequences AA38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX
XX
XX Sequence 287 AA:
SQ
Query Match 76.7%; Score 34.5; DB 20; Length 287;
Best Local Similarity 88.9%; Pred. No. 64;
Matches 8; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 RRF-VHRRP 8
DB 251 rrfavhrrp 259

```

```

XX
XX 09-OCT-1998; 98WO-IB01665.
XX
XX
XX 01-SEP-1998; 98GB-0019016.
XX 06-NOV-1997; 97GB-0023516.
XX 14-NOV-1997; 97GB-0024190.
XX 18-NOV-1997; 97GB-0024386.
XX 27-NOV-1997; 97GB-0025158.
XX 10-DEC-1997; 97GB-0026147.
XX 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Massignani V, Pizsa M, Rappuoli R, Scarlato V;
XX WPI: 1999-327407/27.
XX N-PSDB; AA211984.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 71; 524pp; English.
XX
XX Amino acid sequences AA38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX
XX
XX Sequence 304 AA:
SQ
Query Match 76.7%; Score 34.5; DB 20; Length 304;
Best Local Similarity 88.9%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 RRF-VHRRP 8
DB 251 rrfavhrrp 259

```

```

RESULT 7
AA38512
ID AA38512 standard; Protein: 304 AA.
XX
XX AA38512;
AC
XX
XX 08-OCT-1999 (first entry)
XX
XX Neisseria gonorrhoeae antigenic protein encoded by ORF5.
DE
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
XX Neisseria gonorrhoeae.
OS
XX WO9924578-A2.
XX
XX
XX 20-MAY-1999.
XX
XX

```

```

RESULT 8
AA049814
ID AA049814 standard; Protein: 70 AA.
XX
XX AA049814;
AC
XX
XX 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #10710.
DE
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
OS
XX
XX WO200181581-A2.
XX
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX

```

PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI: 2001-616774/71.  
DR N-PSDB; AAS59545.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1: SEQ ID No 11009; 1069pp; English.  
XX  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 70 AA:  
  
Query Match 75.6%; Score 34; DB 22; Length 70;  
Best Local Similarity 75.0%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 RRFVHRPP 8  
Db 58 rrwshrrp 65  
11:11111  
58 rrwshrrp 65  
  
RESULT 9  
ID AAB75292 standard; Protein: 156 AA.  
XX AAB75292;  
AC AAB75292;  
XX  
DT 03-APR-2001 (first entry)  
XX  
DE Gene 6 human secreted protein homologous amino acid sequence #111.  
XX  
XX Human; immunosuppressive; antiarthritic; antirheumatic; nocotropic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neuroprotective; antibacterial; virucide; fungicide; opthalmological;  
KW vulnery; autoimmune disease; hyperproliferative disorder; cancer;  
KW cardiovascular disorder; cerebrovascular disorder; infection;  
KW nervous system disorder; ocular disorder; chemotaxis; food additive;  
KW secreted protein.  
XX  
XX Homo sapiens.  
XX  
PN WO200077021-A1.  
XX  
PD 21-DEC-2000.  
XX  
XX 01-JUN-2000; 2000WO-US15135.  
XX  
XX 11-JUN-1999; 99US-0138632.  
XX

PA (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX  
PI Rosen CA, Ruben SM, Komatsoulis GA;  
XX WPI: 2001-071257/08.  
XX  
PT Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -  
XX  
XX Disclosure; Page 18-19; 530pp; English.  
XX  
PS This invention relates to polynucleotide sequences AAF63789 - AAF63836  
XX which encode human secreted proteins AAB75260 - AAB75287. Included in the  
CC invention are protein sequences AAB75288 - AAB75341 which are fragments  
CC of the secreted proteins and amino acid sequences with which these  
CC fragments share homology. Examples of the activities of the proteins and  
CC polynucleotides and the activities of their agonists and antagonists  
CC include, immunosuppressive; antiarthritic; antirheumatic;  
CC antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
CC nocotropic; neuroprotective; antibacterial; virucide; fungicide;  
CC opthalmological; and vulnery activity. The protein and polynucleotide  
CC sequences, their agonists and antagonists may be useful for treating,  
CC preventing and diagnosing diseases and disorders such as autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders  
CC e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,  
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,  
CC infections caused by bacteria, viruses and fungi and ocular disorders  
CC e.g. corneal infection. The polypeptides can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. Included in the invention are  
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used  
CC in the identification and characterisation of the DNA and protein  
CC sequences of the invention.  
XX  
SQ Sequence 156 AA:  
  
Query Match 75.6%; Score 34; DB 22; Length 156;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 RRFVHRPP 8  
Db 24 rffvhrkp 31  
111111  
24 rffvhrkp 31  
  
RESULT 10  
ID AAB75293 standard; Protein: 156 AA.  
XX AAB75293;  
AC AAB75293;  
XX  
DT 03-APR-2001 (first entry)  
XX  
DE Human secreted protein sequence encoded by gene 6 SEQ ID NO:112.  
XX  
XX Human; immunosuppressive; antiarthritic; antirheumatic; nocotropic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neuroprotective; antibacterial; virucide; fungicide; opthalmological;  
KW vulnery; autoimmune disease; hyperproliferative disorder; cancer;  
KW cardiovascular disorder; cerebrovascular disorder; infection;  
KW nervous system disorder; ocular disorder; chemotaxis; food additive;  
KW secreted protein.  
XX  
XX Homo sapiens.  
XX  
PN WO200077021-A1.

[illegible]

OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE ) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI: 2001-656860/75.
DR	N-PSDB; ABL09435.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT	
XX	
PS	Disclosure: SEQ ID NO 22788; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins .
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
CC	
XX	
SQ	Sequence 217 AA:
Query Match	75.6%; Score 34; DB 22; Length 217;
Best Local Similarity	87.5%; Pred. No. 61;
Matches	7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY	1 RRFVHRRP 8 
Db	123 rtfvvrtp 130
RESULT 12	
ABBS58399	
ID	ABBS58399 standard; Protein: 243 AA.
XX	
AC	ABBS58399;
XX	
JT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 1989.
XX	
KM	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
FN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE ) PE CORP NY.



XX Venter JC, Adams M, Li FMD, Myers EM;  
XX WPI: 2001-656860/75.  
DR N-PSDB; ABL02502.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 1989; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABB57737-ABB72072).  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 243 AA;

Query Match 75.6%; Score 34; DB 22; Length 243;  
Best Local Similarity 75.0%; Pred. No. 68;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RRFVHRRP 8  
| |||: ||  
Db 104 rffvhrkp 111

RESULT 13  
ABG18986  
ID ABG18986 standard; Protein; 273 AA.  
XX  
AC ABG18986;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #18977.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS83173.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 49345; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 273 AA;

Query Match 75.6%; Score 34; DB 22; Length 273;  
Best Local Similarity 75.0%; Pred. No. 76;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RRFVHRRP 8  
| |||: ||  
Db 125 rffvhrkp 132

RESULT 14  
ABG10923  
ID ABG10923 standard; Protein; 280 AA.  
XX  
AC ABG10923;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10914.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS75110.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 41282; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
CC  
SQ Sequence 280 AA;

Query Match 75.6%; Score 34; DB 22; Length 280;  
Best Local Similarity 75.0%; Pred. No. 78;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVHRPP 8  
| | | | | | | |  
Db 125 rffvhrkp 132

RESULT 15  
ABG10922  
ID ABG10922 standard; Protein; 282 AA.  
XX  
AC ABG10922;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10913.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB: AAS75109.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 41281; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
CC  
SQ Sequence 282 AA;

Query Match 75.6%; Score 34; DB 22; Length 282;  
Best Local Similarity 75.0%; Pred. No. 79;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVHRPP 8  
| | | | | | | |  
Db 104 rffvhrkp 111

Search completed: August 15, 2002, 11:49:03  
Job time: 559 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 11:49:35 ; Search time 27.13 Seconds  
(without alignments)  
7.203 Million cell updates/sec

Title: US-09-613-092a-7\_COPY\_3\_10  
Perfect score: 45  
Sequence: 1 RRFVHRRP 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	75.6	244	2	US-08-933-750C-23
2	34	75.6	244	4	US-09-234-613-23
3	33	73.3	155	2	US-08-670-186-2
4	31	68.9	282	2	US-08-997-080-134
5	31	68.9	282	2	US-08-997-362-134
6	31	68.9	282	4	US-09-085-855-134
7	31	68.9	282	4	US-09-324-542-134
8	31	68.9	297	2	US-08-997-080-177
9	31	68.9	297	2	US-08-997-362-177
10	31	68.9	297	4	US-09-095-855-177
11	31	68.9	297	4	US-09-324-542-177
12	31	68.9	315	3	US-08-558-135-5
13	31	68.9	455	2	US-08-997-080-121
14	31	68.9	455	2	US-08-997-362-121
15	31	68.9	455	4	US-09-095-855-121
16	31	68.9	455	4	US-09-324-542-121
17	31	68.9	471	2	US-08-997-080-172
18	31	68.9	471	2	US-08-997-362-172
19	31	68.9	471	4	US-09-095-855-172
20	31	68.9	471	4	US-09-324-542-172
21	31	68.9	591	2	US-08-889-402-1
22	31	68.9	605	2	US-08-889-402-2
23	31	68.9	2890	4	US-09-413-814-67
24	31	68.9	3200	2	US-08-477-451-8
25	31	68.9	3798	3	US-09-335-409-6
26	31	68.9	3798	4	US-09-368-102-6
27	31	68.9	3798	4	US-09-567-969-6

28	31	68.9	3798	4	US-09-568-480-6	Sequence 6, Appl1
29	31	68.9	3798	4	US-09-568-486-6	Sequence 6, Appl1
30	31	68.9	3798	4	US-09-568-472-6	Sequence 6, Appl1
31	30	66.7	278	2	US-08-701-191A-39	Sequence 39, Appl1
32	30	66.7	304	2	US-08-701-191A-30	Sequence 30, Appl1
33	30	66.7	392	4	US-09-147-926-2	Sequence 2, Appl1
34	30	66.7	530	5	PCT-US95-08493-2	Sequence 2, Appl1
35	30	66.7	860	5	PCT-US95-08493-19	Sequence 19, Appl1
36	30	66.7	868	1	US-08-374-834-1	Sequence 1, Appl1
37	30	66.7	868	2	US-08-644-271-1	Sequence 1, Appl1
38	30	66.7	868	5	PCT-US95-08493-21	Sequence 21, Appl1
39	30	66.7	869	1	US-08-374-834-16	Sequence 16, Appl1
40	30	66.7	869	2	US-08-644-271-29	Sequence 29, Appl1
41	30	66.7	946	5	PCT-US95-08493-13	Sequence 13, Appl1
42	30	66.7	1052	3	US-08-863-118-1	Sequence 1, Appl1
43	30	66.7	1052	3	US-08-863-118-2	Sequence 2, Appl1
44	30	66.7	1052	4	US-09-377-310-2	Sequence 2, Appl1
45	30	66.7	1053	3	US-08-863-118-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-08-933-750C-23  
: Sequence 23, Application US/08933750C  
: Patent No. 5932442  
GENERAL INFORMATION:  
: APPLICANT: Lal, Preeti  
: APPLICANT: Hillman, Jennifer L.  
: APPLICANT: Bandman, Olga  
: APPLICANT: Shah, Purvi  
: APPLICANT: Au-Young, Janice  
: APPLICANT: Yue, Henry  
: APPLICANT: Guegler, Karl J.  
: APPLICANT: Corley, Neil C.  
: TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
: NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Incyte Pharmaceuticals, Inc.  
: STREET: 3174 Porter Drive  
: CITY: Palo Alto  
: STATE: CA  
: COUNTRY: USA  
: ZIP: 94304  
COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: DOS  
: SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/933,750C  
: FILING DATE: September 23, 1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
: NAME: Billings, Lucy J.  
: REGISTRATION NUMBER: 36,749  
: REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 415-855-0555  
: TELEFAX: 415-845-4166  
TELEX:  
: INFORMATION FOR SEQ ID NO: 23:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 244 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: IMMEDIATE SOURCE:  
: LIBRARY: SPLN0704

CLONE: 1561587  
US-08-933-750C-23

Query Match 75.6%; Score 34; DB 2; Length 244;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRFVHRRP 8  
1 |||:|  
DB 105 RRFVHRRP 112

## RESULT 2

US-09-234-613-23  
; Sequence 23, Application US/09234613  
; Patent No. 6132973  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/234,613  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750  
; FILING DATE: September 23, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 244 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: SPLNOT04  
; CLONE: 1561587  
; US-09-234-613-23

Query Match 75.6%; Score 34; DB 4; Length 244;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RRFVHRRP 8  
1 |||:|

DB 105 RRFVHRRP 112

## RESULT 3

US-08-670-186-2  
; Sequence 2, Application US/08670186  
; Patent No. 585343  
; GENERAL INFORMATION:  
; APPLICANT: SUN, SAMUEL S.M.  
; APPLICANT: XIONG, LIWEN  
; APPLICANT: HU, ZHONG  
; APPLICANT: CHEN, HANG  
; TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FORSTER  
; STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/670,186  
; FILING DATE: 21-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MORASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 23461-20007.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 822-0168  
; TELEX: 90-4030 MRSNFORSMH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-670-186-2

Query Match 73.3%; Score 33; DB 2; Length 155;  
Best Local Similarity 71.4%; Pred. No. 24;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRFVHRR 7  
1 |||:|  
DB 54 GRFHRR 60

## RESULT 4

US-08-997-080-134  
; Sequence 134, Application US/08997080  
; Patent No. 5968524  
; GENERAL INFORMATION:  
; APPLICANT: WATSON, JAMES D.  
; APPLICANT: TAN, PAUL L.J.  
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-  
; NUMBER OF SEQUENCES: 194  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4105  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,080  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 282 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-997-080-134

Query Match 68.9%; Score 31; DB 2; Length 282;  
Best Local Similarity 71.4%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 REVHRP 8  
Db 267 RLVHRKP 273

RESULT 5  
US-08-997-362-134  
Sequence 134, Application US/08997362  
Patent No. 5985287  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Miyama, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,362  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970  
FILING DATE: June 12, 1997  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347  
FILING DATE: August 29, 1996

ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 282 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-997-362-134

Query Match 68.9%; Score 31; DB 2; Length 282;  
Best Local Similarity 71.4%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 REVHRP 8  
Db 267 RLVHRKP 273

RESULT 6  
US-09-095-855-134  
Sequence 134, Application US/09095855  
Patent No. 6160093  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Compounds and Methods for  
Treatment and Diagnosis of Mycobacterial Infections  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970  
FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362  
FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 282 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-095-855-134

Query Match 68.9%; Score 31; DB 4; Length 282;  
Best Local Similarity 71.4%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RYVHRP 8  
DB 267 RLVHRP 273

RESULT 7  
US-09-324-542-134  
Sequence 134, Application US/09324542  
Patent No. 6328978  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Tan, Paul L.J.  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Methods and Compounds for the Treatment  
FILE REFERENCE: 11000.1007c1  
CURRENT APPLICATION NUMBER: US/09/324,542  
CURRENT FILING DATE: 1999-06-02  
EARLIER APPLICATION NUMBER: US 08/997,080  
EARLIER FILING DATE: 1997-12-23  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 134  
LENGTH: 282  
TYPE: PRT  
ORGANISM: Mycobacterium vaccae  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (145)..(145)  
NAME/KEY: UNSURE  
LOCATION: (151)..(151)  
US-09-324-542-134

Query Match 68.9%; Score 31; DB 4; Length 282;  
Best Local Similarity 71.4%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RYVHRP 8  
DB 267 RLVHRP 273

RESULT 8  
US-08-997-080-177  
Sequence 177, Application US/08997080  
Patent No. 5968524  
GENERAL INFORMATION:  
APPLICANT: WATSON, JAMES D.  
APPLICANT: TAN, PAUL L.J.  
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,080  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 177:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-997-080-177

Query Match 68.9%; Score 31; DB 2; Length 297;  
Best Local Similarity 71.4%; Pred. No. 1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RYVHRP 8  
DB 267 RLVHRP 273

RESULT 9  
US-08-997-362-177  
Sequence 177, Application US/08997362  
Patent No. 5968287  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Hiyyama, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,362  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. Patent Application No. 5965287 08/873,970  
FILING DATE: June 12, 1997  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347  
FILING DATE: August 29, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007



REFERENCE/DOCKET NUMBER: 11000.1002c2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 177:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-997-362-177

Query Match  
Best Local Similarity 68.9%; Score 31; DB 2; Length 297;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 REVHRRP 8  
Db 267 RLVHRRP 273

RESULT 10  
US-09-095-855-177  
Sequence 177, Application US/09095855  
Patent No. 6160093  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Compounds and Methods for  
TREATMENT OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970  
FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362  
FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 177:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-095-855-177

Query Match  
Best Local Similarity 68.9%; Score 31; DB 4; Length 297;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 REVHRRP 8  
Db 267 RLVHRRP 273

RESULT 11  
US-09-324-542-177  
Sequence 177, Application US/09324542  
Patent No. 6328978  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Tan, Paul L.J.  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Methods and Compounds for the Treatment  
of Immunologically-Mediated Skin Disorders  
FILE REFERENCE: 11000.1007c1  
CURRENT APPLICATION NUMBER: US/09/324,542  
EARLIER FILING DATE: 1999-06-02  
EARLIER APPLICATION NUMBER: US 08/997,080  
EARLIER FILING DATE: 1997-12-23  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 177  
LENGTH: 297  
TYPE: PRT  
ORGANISM: Mycobacterium vaccae  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (145)...(145)  
NAME/KEY: UNSURE  
LOCATION: (151)...(151)  
US-09-324-542-177

Query Match  
Best Local Similarity 68.9%; Score 31; DB 4; Length 297;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 REVHRRP 8  
Db 267 RLVHRRP 273

RESULT 12  
US-08-558-135-5  
Sequence 5, Application US/08558135  
Patent No. 6090631  
GENERAL INFORMATION:  
APPLICANT: Catterall, William A.  
APPLICANT: Sheng, Zu-Hang  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING  
FOR PRESYNAPTIC CALCIUM CHANNEL BLOCKERS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/558,135

FILING DATE: 13-NOV-1995  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.602C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 315 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-558\*135-5

Query Match 68.9%; Score 31; DB 3; Length 315;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRVHRRP 8  
Db 242 RRVHRRP 249

RESULT 13  
US-08-997-080-121  
Sequence 121, Application US/08997080  
Patent No. 5968524  
GENERAL INFORMATION:  
APPLICANT: WATSON, JAMES D.  
APPLICANT: TAN, PAUL L.J.  
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,080  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 455 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-997-080-121

Query Match 68.9%; Score 31; DB 2; Length 455;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RRVHRRP 8  
Db 440 RRVHRRP 446

RESULT 14  
US-08-997-362-121  
Sequence 121, Application US/08997362  
Patent No. 5985287  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Hiyama, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,362  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970  
FILING DATE: June 12, 1997  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347  
FILING DATE: August 29, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 455 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-997-362-121

Query Match 68.9%; Score 31; DB 2; Length 455;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RRVHRRP 8  
Db 440 RRVHRRP 446

RESULT 15  
US-09-095-855-121

; Sequence 121, Application US/09095855  
; Patent No. 6160093  
; GENERAL INFORMATION:  
; APPLICANT: Tan, Paul  
; APPLICANT: Visser, Elizabeth  
; APPLICANT: Skinner, Margot  
; APPLICANT: Prestidge, Ross  
; TITLE OF INVENTION: Compounds and Methods for  
; TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections  
; NUMBER OF SEQUENCES: 208  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/095,855  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/705,347  
; FILING DATE: 29-AUG-1996  
; APPLICATION NUMBER: 08/873,970  
; FILING DATE: 12-JUN-1997  
; APPLICATION NUMBER: 08/997,362  
; FILING DATE: 23-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sleath, Janet  
; REGISTRATION NUMBER: 37,007  
; REFERENCE/DOCKET NUMBER: 11000.1002c3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-269-0565  
; TELEFAX: 206-269-0563  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 121:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 455 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-095-855-121

Query Match 68.9%; Score 31; DB 4; Length 455;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RFVHRP 8  
| | | | |  
Db 440 RLVHRKP 446

Search completed: August 15, 2002, 11:49:36  
Job time: 277 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:50:18 ; Search time 35.29 Seconds  
(Without alignments)  
21.783 Million cell updates/sec

Title: US-09-613-092a-7\_COPY\_3\_10  
Perfect score: 45  
Sequence: 1 RRFVHRR 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	84.4	434	2 F75425	tRNA nucleotidyltr
2	35	77.8	72	2 T12874	hypothetical prote
3	35	77.8	246	2 AG2988	competence protein
4	35	77.8	291	2 A99295	compeptence protein
5	35	77.8	403	2 B70961	probable esterase
6	34	75.6	80	2 G81065	hypothetical prote
7	34	75.6	248	2 S30194	ribosomal protein
8	33	73.3	33	2 S28842	mabulin II chain
9	33	73.3	118	2 T35747	hypothetical prote
10	33	73.3	155	2 UC5379	mabulin II precur
11	33	73.3	246	2 T47398	hypothetical prote
12	33	73.3	301	2 AC0266	oligopeptide trans
13	33	73.3	345	1 S36607	protein kinase (EC
14	33	73.3	496	2 T50146	probable era/thbf
15	33	73.3	548	2 T30369	DNA ligase-like pr
16	33	73.3	577	2 F72122	hypothetical prote
17	33	73.3	577	2 C86499	hypothetical prote
18	33	73.3	722	2 C88397	protein H04J21.3 f
19	33	73.3	786	2 T49414	related to ahmpl p
20	32	71.1	138	2 H72534	hypothetical prote
21	32	71.1	140	2 E83381	probable ring-clea
22	32	71.1	202	2 T01601	probable PREG1-lik
23	32	71.1	290	2 AH3448	xbhc protein (assi
24	32	71.1	295	2 F70936	hypothetical prote
25	32	71.1	370	2 T46150	protein kinase ATN
26	32	71.1	382	2 AE0681	probable ABC trans
27	32	71.1	451	2 A70539	hypothetical prote
28	32	71.1	527	2 T37055	probable oxidoredu
29	32	71.1	776	2 T29064	hyaluronate lyase

30	32	71.1	1263	2 T15496	hypothetical prote
31	32	71.1	1527	1 RNZMB2	DNA-directed RNA p
32	32	71.1	1792	2 T13939	myosin V - fruit f
33	31	68.9	193	2 A82705	oligotribonuclease
34	31	68.9	193	2 A87333	hypothetical prote
35	31	68.9	202	2 B86447	hypothetical prote
36	31	68.9	238	2 T47386	hypothetical prote
37	31	68.9	320	2 B90325	conserved hypobhet
38	31	68.9	360	2 A85016	hypothetical prote
39	31	68.9	382	2 H83672	hypothetical prote
40	31	68.9	428	2 A89900	conserved hypobhet
41	31	68.9	444	2 T01721	hypothetical prote
42	31	68.9	444	2 T15907	hypothetical prote
43	31	68.9	446	2 S46802	hypothetical prote
44	31	68.9	557	2 A55933	paxillin - human
45	31	68.9	634	2 JC2376	dnak-type molecula

## ALIGNMENTS

RESULT 1  
F75425  
tRNA nucleotidyltransferase/poly A polymerase family protein - Deinococcus radioduran  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Aug-2000  
C:Accession: F75425  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
S.; Shen, M.; Vamathevan, J.V.; Lam, P.; McDonald, L.; Ulterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: F75425  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-434 <WHI>  
A:Cross-references: GB:AE001968; GB:AE000513; NID:96458930; PIDN:AAF10763.1; PID:9645  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1191  
A:Map position: 1  
C:Superfamily: Escherichia coli tRNA adenylyltransferase

Query Match 84.4% Score 38; DB 2; Length 434;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
DB 322 RRFVHRR 328

RESULT 2  
T12874  
hypothetical protein yocN - Bacillus subtilis phage SPBc2  
C:Species: Bacillus subtilis phage SPBc2  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 15-Oct-1999  
R:Azarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.  
submitted to the EMBL Data Library, August 1997  
A:Description: The complete nucleotide sequence of the Bacillus subtilis spBc2 pro  
A:Reference number: Z17583  
A:Accession: T12874  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-72 <LAZ>  
A:Cross-references: EMBL:AF020713; NID:93025478; PID:93025588; PIDN:AA013083.1  
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Ehtian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hildbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Muehl, Y.; M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Walters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MID:198044033  
A:Accession: F69920  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-72 <KUN>  
A:Cross-references: GB:299114; GB:AL009126; NID:g2634230; PIDN:CAB13949.1; PID:EL185529;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yocN  
A:Note: yocN

Query Match 77.8%; Score 35; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FVHRRP 8  
|||||  
DB 60 FVHRRP 65

RESULT 3  
AG2988  
C:Species: *Agrobacterium tumefaciens* (strain C58, Dupont)  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AG2988  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AG2988  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <KUR>  
A:Cross-references: GB:AE008609; PIDN:AAL44325.1; PID:g17741916; GSPDB:GNO0187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: comF  
A:Map position: linear chromosome  
C:Superfamily: transformation competence-related protein comF

Query Match 77.8%; Score 35; DB 2; Length 246;  
Best Local Similarity 85.7%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
|||||  
DB 135 RRFVHRR 141

RESULT 4  
A9295  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: A9295

R;Goddeer, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium*  
A:Reference number: A97359; PMID:11743194  
A:Accession: A9295  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-291 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK89883.1; PID:g15159828; GSPDB:GNO0170  
C:Genetics:  
A:Gene: AGR\_L\_2633  
A:Map position: linear chromosome

Query Match 77.8%; Score 35; DB 2; Length 291;  
Best Local Similarity 85.7%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
|||||  
DB 180 RRFVHRR 186

RESULT 5  
B70961  
C:Species: *Mycobacterium tuberculosis* (strain H37RV)  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: B70961  
R:Colet, S.T.; Brosch, R.; Parhail, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Seares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
A:Reference number: A70500; MID:98295987  
A:Accession: B70961  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-403 <COL>  
A:Cross-references: GB:292669; GB:AL123456; NID:g3242271; PIDN:CAB07015.1; PID:g18715  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: lipC  
C:Superfamily: *Mycobacterium tuberculosis* probable esterase

Query Match 77.8%; Score 35; DB 2; Length 403;  
Best Local Similarity 85.7%; Pred. No. 33;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
|||||  
DB 125 RRFVHRR 131

RESULT 6  
G81065  
C:Species: *Neisseria meningitidis* (strain MC58 serogro  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: G81065  
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;  
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A:Reference number: A81000; MID:20175755  
A:Accession: G81065  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-80 <TEF>  
A:Cross-references: GB:AE002509; GB:AE002098; NID:g7226832; PID:AAE41942.1; PID:g722683  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1589

Query Match 75.6%; Score 34; DB 2; Length 80;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
Db 63 RQFVHRR 69

RESULT 7  
S30194  
ribosomal protein S6, cytosolic - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jul-2000  
C:Accession: S30194; A47240  
R:Spencer, T.A.; Mackie, G.A.  
Biochim. Biophys. Acta 1172, 332-334, 1993  
A:Title: The nucleotide sequence of a cloned cDNA encoding ribosomal protein S6 from Drc  
A:Reference number: S30194; MUID:93192329  
A:Accession: S30194  
A:Molecule type: mRNA  
A:Residues: 1-248 <SPE>  
A:Cross-references: EMBL:L07881; NID:g158335; PID:AAA28871.1; PID:g158336  
R:Watson, K.L.; Konrad, K.D.; Woods, D.F.; Bryant, P.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 11302-11306, 1992  
A:Title: Drosophila homolog of the human S6 ribosomal protein is required for tumor supp  
A:Reference number: A47240; MUID:93087515  
A:Accession: A47240  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-248 <MAT>  
A:Cross-references: GB:L01558; NID:g3540130; PID:AAAC34306.1; PID:g158338  
A:Note: sequence extracted from NCBI backbone (NCBIR:119775)  
C:Genetics:  
A:Gene: FLYBase:RPS6  
A:Cross-references: FLYBase:FBgn0004922  
C:Superfamily: rat ribosomal protein S6  
C:Keywords: protein biosynthesis; ribosome

Query Match 75.6%; Score 34; DB 2; Length 248;  
Best Local Similarity 87.5%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVHRR 8  
Db 154 RRFVHRR 161

RESULT 8  
S28842  
mabinlin II chain A - Yunnan caper  
C:Species: Capparis masakal (Yunnan caper)  
C>Date: 22-Jan-1994 #sequence\_revision 27-Oct-1995 #text\_change 25-Oct-1996  
C:Accession: S28842  
R:Li, X.; Maeda, S.; Hu, Z.; Aluchi, T.; Nakaya, K.; Kurihara, Y.  
Eur. J. Biochem. 211, 281-287, 1993  
A:Title: Purification, complete amino acid sequence and structural characterization of t  
A:Reference number: S28842; MUID:95145958  
A:Accession: S28842  
A:Molecule type: protein  
A:Residues: 1-33 <LID>  
C:Superfamily: wheat alpha-amylase inhibitor  
C:Keywords: sweet taste

Query Match 73.3%; Score 33; DB 2; Length 33;  
Best Local Similarity 71.4%; Pred. No. 7.2;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
Db 19 QRFVHRR 25

RESULT 9  
T35747  
hypothetical protein SC7H2.16c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35747  
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream,  
submitted to the EMBL Data Library, August 1999  
A:Reference number: 221598  
A:Accession: T35747  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-118 <SAU>  
A:Cross-references: EMBL:AL109732; PID:CA852058.1; GSPDB:GN00070; SCODEB:SC7H2.16c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCODEB:SC7H2.16c

Query Match 73.3%; Score 33; DB 2; Length 118;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRFVHRR 8  
Db 53 RFLVHRR 60

RESULT 10  
JC5379  
mabinlin II precursor - Yunnan caper  
C:Species: Capparis masakal (Yunnan caper)  
C>Date: 02-Jun-1997 #sequence\_revision 12-Sep-1997 #text\_change 20-Jun-2000  
C:Accession: JC5379; PC4316  
R:Nirasawa, S.; Masuda, Y.; Nakaya, K.; Kurihara, Y.  
Gene 181, 225-227, 1996  
A:Title: Cloning and sequencing of a cDNA encoding a heat-stable sweet protein, mabln  
A:Reference number: JC5379; MUID:97128796  
A:Accession: JC5379  
A:Molecule type: mRNA  
A:Residues: 1-155 <NIR1>  
A:Cross-references: DDBJ:D83997; NID:g1817545; PID:BAI2204.1; PID:g1817546  
A:Accession: PC4316  
A:Molecule type: protein  
A:Residues: 36-41; 149-154 <NIR2>  
A:Experimental source: seed  
C:Superfamily: wheat alpha-amylase inhibitor  
C:Keywords: sweet taste  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-35/Domain: amino-terminal propeptide #status predicted <PRO>  
F:36-68/83-154/Product: mabinlin II #status experimental <MAT>  
F:36-68/Domain: mabinlin II A chain #status experimental <CHA>  
F:69-83/Domain: mabinlin linker peptide #status predicted <LNK>  
F:83-154/Domain: mabinlin II B chain #status experimental <CHR>

Query Match 73.3%; Score 33; DB 2; Length 155;  
Best Local Similarity 71.4%; Pred. No. 32;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
Db 54 QRFVHRR 60

RESULT 11  
T47398  
hypothetical protein T18D12.130 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47398  
R:Massenet, O.; Clabault, G.; Quigley, F.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24464  
A:Accession: T47398  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <MAS>  
A:Cross-References: EMBL:AL138644  
A:Experimental source: cultivar Columbia; BAC clone T18D12  
C:Genetics:  
A:Map position: 3  
A:Introns: 48/2; 73/1; 145/1; 189/1  
A>Note: T18D12.130

Query Match 73.3%; Score 33; DB 2; Length 246;  
Best Local Similarity 85.7%; Pred. No. 50;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 RRFVHRR 8  
DB 208 RFDHRRP 214

RESULT 12  
AC0266  
oligopeptide transport system permease protein oppc [imported] - Yersinia pestis (strain  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AC0266  
R:Parhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
demo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AC0266  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-301 <KUR>  
A:Cross-References: GB:AL590842; PIDN:CAC90991.1; PID:g15980186; GSPDB:GN00175  
C:Genetics:  
A:Gene: oppc  
C:Superfamily: oligopeptide permease protein oppb

Query Match 73.3%; Score 33; DB 2; Length 301;  
Best Local Similarity 85.7%; Pred. No. 60;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRFVHRR 7  
DB 30 RRFVHNR 36

RESULT 13  
S36607  
protein kinase (EC 2.7.1.37) crk2 [similarity] - Trypanosoma brucei  
C:Species: Trypanosoma brucei  
C>Date: 16-Jun-2000 #sequence\_revision 16-Jun-2000 #text\_change 16-Jun-2000  
C:Accession: S36607  
R:Smith, G.; Motttram, J.  
submitted to the EMBL Data Library, August 1993  
A:Description: A family of CDC2-related kinases in Trypanosoma brucei.  
A:Reference number: S36607  
A:Accession: S36607

A:Molecule type: DNA  
A:Residues: 1-345 <SMI>  
A:Cross-References: EMBL:X74598; NID:g9397161; PIDN:CAA52676.1; PID:g9397162  
A:Experimental source: strain Istat  
C:Genetics:  
A:Gene: crk2  
C:Superfamily: kinase-related transforming protein; protein kinase homology  
C:Keywords: ATP; phosphotransferase  
F:44-297/Domain: protein kinase homology <KIN>  
F:52-60/Region: protein kinase ATP-binding motif  
F:75,93,168,170/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 73.3%; Score 33; DB 1; Length 345;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRFVHR 6  
DB 162 RRFVHRP 167

RESULT 14  
T50146  
probable era/thof family GTP-binding protein [imported] - fission yeast (Schizosaccha  
C:Species: Schizosaccharomyces pombe  
C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T50146  
R:Hamlin, N.; Churcher, C.M.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, November 1999  
A:Reference number: Z25043  
A:Accession: T50146  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-496 <RAM>  
A:Cross-References: EMBL:AL132798; PIDN:CAB60697.1; GSPDB:GN00066; SPDB:SPAC222.05c  
A:Experimental source: strain 972h(-); cosmid c222  
C:Genetics:  
A:Gene: SPDB:SPAC222.05c  
A:Map position: 1  
A:Introns: 50/3  
C:Superfamily: thiophen / furan oxidation protein; translation elongation factor Tu h

Query Match 73.3%; Score 33; DB 2; Length 496;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRFVHRR 8  
DB 17 RRFVHQIP 24

RESULT 15  
T30369  
DNA ligase-like protein - Lymantria dispar nuclear polyhedrosis virus  
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdNPV  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T30369  
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; R  
Virology 253, 17-34, 1999  
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria  
A:Reference number: Z20836; MUID:99124785  
A:Accession: T30369  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-548 <KUZ>  
A:Cross-References: EMBL:AP081810; NID:g3822234; PIDN:AAC70207.1; PID:g3822236  
C:Superfamily: vaccinia virus DNA ligase

Query Match 73.3%; Score 33; DB 2; Length 548;  
Best Local Similarity 100.0%; Pred. No. 11e+02;



Matches	6;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	RRVYHR	6						
Db	333	RRVYHR	338						

Search completed: August 15, 2002, 11:50:19  
Job time: 245 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:58:17 ; Search time 18.61 Seconds

(Without alignments)  
16.645 Million cell updates/sec

Title: US-09-613-092a-7\_COPY\_3\_10  
Perfect score: 45  
Sequence: 1 RRFVHRP 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	34	75.6	248	1 RS6_DROME
2	33	73.3	155	1 ZSS2_CAPPA
3	33	73.3	345	1 CC22_TRYBB
4	33	73.3	496	1 MSS1_SCHPO
5	32	71.1	140	1 YK24_PSEAE
6	32	71.1	263	1 VCY3_ASTLO
7	32	71.1	451	1 YB28_MYCTU
8	32	71.1	551	1 RM56_MOUSE
9	32	71.1	1479	1 RPOD_MHEAT
10	32	71.1	1527	1 RPOD_MAIZE
11	31	68.9	193	1 ORN_XYLFA
12	31	68.9	421	1 WCOM_ADECC
13	31	68.9	421	1 WCOM_ADECC
14	31	68.9	446	1 YHG4_YEAST
15	31	68.9	557	1 PAXI_HUMAN
16	31	68.9	634	1 DNK2_SYNP7
17	31	68.9	898	1 NTA_BETVE
18	31	68.9	1513	1 RPOD_ORYSA
19	31	68.9	1997	1 OTOF_HUMAN
20	31	68.9	1997	1 OTOF_MOUSE
21	31	68.9	2067	1 BIMB_EMENT
22	31	68.9	2424	1 CCNA_RABIT
23	30	66.7	125	1 YN25_CAEEL
24	30	66.7	176	1 EMT2_RAT
25	30	66.7	269	1 RFAZ_SALTY
26	30	66.7	302	1 OPFC_ECOLI
27	30	66.7	302	1 OPFC_SALTY
28	30	66.7	342	1 AOX_HANAN
29	30	66.7	349	1 ALEI_ECOLI
30	30	66.7	541	1 BBRI_SCHOL
31	30	66.7	637	1 PRIM_LACIA
32	30	66.7	704	1 GUGB_YEAST
33	30	66.7	757	1 HT16_HYDAT

34	30	66.7	1052	1 FAK1_HUMAN	005397 homo sapien
35	30	66.7	1052	1 FAK1_MOUSE	P4152 mus musculu
36	30	66.7	1053	1 FAK1_CHICK	000944 gallus gall
37	30	66.7	1055	1 FAK1_RAT	035346 rattus norv
38	30	66.7	1063	1 PDRL_YEAST	P12383 saccharomyc
39	30	66.7	1068	1 FAK1_XENLA	Q91738 xenopus lae
40	30	66.7	1071	1 TRI_THEAC	P96086 thermoplasm
41	29	64.4	205	1 RS2_AERPE	Q9YB45 aeropyrium p
42	29	64.4	217	1 RL3_MYCHO	006044 mycobacteri
43	29	64.4	217	1 RL3_MYCHO	P30762 mycobacteri
44	29	64.4	217	1 RL3_MYCHO	P5049 mycobacteri
45	29	64.4	370	1 MDS1_YEAST	P38615 saccharomyc

## ALIGNMENTS

RESULT 1

ID	RS6_DROME	STANDARD:	PRT:	248 AA.
AC	P29327: Q94993; Q9W3N4;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	40S ribosomal protein S6.			
GN	RP56 OR L(1)A1R8 OR HEN OR CG10944.			
OS	Drosophila melanogaster (fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93087515; PubMed=1454811;			
RA	Watson K.U., Konrad K.D., Woods D.F., Bryant P.J.;			
RT	"Drosophila homolog of the human S6 ribosomal protein is required for			
RT	tumor suppression in the hematopoietic system.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:11302-11306(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93192327; PubMed=8448211;			
RA	Spencer T.A., Mackie G.A.;			
RT	"The nucleotide sequence of a cloned cDNA encoding ribosomal protein			
RT	S6 from Drosophila melanogaster.";			
RL	Biochim. Biophys. Acta 1172:332-334(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.			
RX	MEDLINE=94018587; PubMed=8412647;			
RA	Stewart M.J., Denell R.;			
RT	"The Drosophila ribosomal protein S6 gene includes a 3' triplication			
RT	that arose by unequal crossing-over.";			
RL	Mol. Biol. Evol. 10:1041-1047(1993).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM A).			
RC	STRAIN-BERKELEY;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazek R.G., Chang M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Barker E.G., Helt G., Nelson C.R., Mkilios G.L.G.,			
RA	Abriil J.F., Ayguyan A., An H.-J., Andrews-Pfankoch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Botkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,			
RA	Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Dudin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			



FT CONFLICT 148 148 A -> T (IN REF. 1).  
 FT CONFLICT 153 153 A -> T (IN REF. 1).  
 SO SEQUENCE 155 AA; 18089 MM; 72E885DEDC2D46A CRC64;

Query Match 73.3%; Score 33; DB 1; Length 155;  
 Best Local Similarity 71.4%; Pred. No. 8.4;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRFVHRR 7  
 Db 54 QRFTHRR 60

## RESULT 3

CC22\_TRYBB STANDARD; PRT; 345 AA.  
 AC P54665;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-1999 (Rel. 38, Last annotation update)  
 DE Cell division control protein 2 homolog 2 (EC 2.7.1.-).  
 GN CKR2.  
 OS Trypanosoma brucei brucei.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ISTAT;  
 RX MEDLINE=96009893; PubMed=7557404;  
 RA Mottiram J., Smith G.;  
 RT "A family of trypanosome cdc2-related protein kinases."  
 RL Gene 162.147-152(1995).  
 CC -1- FUNCTION: PROBABLY INVOLVED IN THE CONTROL OF THE CELL CYCLE.  
 CC -1- ENZYME REGULATION: PHOSPHORYLATION AT SER-56 OR TYR-57 INACTIVATES THE ENZYME (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS A STABLE BUT NON-COVALENT COMPLEX WITH A REGULATORY SUBUNIT AND WITH A CYCLIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SRR/THR FAMILY OF PROTEIN KINASES.  
 CC -1- CDDX SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X74598; CAAS2676.1; -.  
 DR HSSP; P24941; ICKP.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_Pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KM Phosphorylation.  
 FT DOMAIN 46 328 PROTEIN KINASE.  
 FT NP\_BIND 52 60 ATP (BY SIMILARITY).  
 FT BINDING 75 75 ATP (BY SIMILARITY).  
 FT ACT\_SITE 168 168 BY SIMILARITY.  
 FT MOD\_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 57 57 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 345 AA; 39238 MM; 6E4E88D2571931B8 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRFVHR 6  
 Db 162 RRFVHR 167

## RESULT 4

MSL1\_SCHPO STANDARD; PRT; 496 AA.  
 AC Q9PTE7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GTPase MSL1 homolog, mitochondrial precursor.  
 GN SPAC222.05C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Hamlin N., Churcher C.M., McDougall R.C., Rajandream M.A.,  
 RA Barrell B.G.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -1- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING PROTEINS. TRME SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL132798; CAB60697.1; -.  
 DR InterPro: IPR002917; MMR\_HSR1.  
 DR Pfam; PF01926; MMR\_HSR1; 1.  
 KW tRNA processing; Mitochondrion; GTP-binding; Transit peptide.  
 FT TRANST 1 ? MITOCHONDRION (POTENTIAL).  
 FT CHAIN ? 496 GTPASE MSL1 HOMOLOG.  
 FT NP\_BIND 246 253 GTP (POTENTIAL).  
 SO SEQUENCE 496 AA; 55315 MM; 092F7E1E1CDD8002 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 496;  
 Best Local Similarity 75.0%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RRFVHRR 8  
 Db 17 RRFVHQP 24

## RESULT 5

YK24\_PSEAE STANDARD; PRT; 140 AA.  
 ID YK24\_PSEAE  
 AC P23205; Q91290;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein PA2024.  
 GN PA2024.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Binkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garner R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 RN [2]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RC STRAIN=PA08;  
 RX MEDLINE=91194546; PubMed=1849605;  
 RA Perry A.C.F., Ni Bhriain N., Brown N.L., Rouch D.A.;  
 RT "Molecular characterization of the gor gene encoding glutathione  
 RT reductase from *Pseudomonas aeruginosa*: determinants of substrate  
 RT specificity among pyridine nucleotide-disulphide oxidoreductases.";  
 RL Mol. Microbiol. 5:163-171(1991).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE004629; AAG05412.1; -  
 DR EMBL: X54201; CAA38121.1; -  
 DR PIR: S15235; S15235.  
 DR InterPro: IPR000325; Glyoxalase\_1.  
 DR Pfam: PF00903; Glyoxalase\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 140 AA; 15373 MW; 26221B19B832BF38 CRC64;

OY 2 REVHRRP 8  
 1 :|||||  
 Db 29 RVTHRRP 35

Query Match 71.1%; Score 32; DB 1; Length 140;  
 Best Local Similarity 71.4%; Pred. No. 12;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 6  
 YCY3\_ASTLO STANDARD; PRT; 263 AA.  
 ID YCY3\_ASTLO  
 AC P58147;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 32.5 kDa protein in rpl14-rpl12 intergenic region  
 DE (ORF263).  
 OS *Astasia longa* (Euglenophyceae algae).  
 OG Chloroplast.  
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Astasia.  
 OX NCBI\_TaxID=3037;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CCAP 1204-17A;  
 RX MEDLINE=21080522; PubMed=11212895;  
 RA Gockel G., Hachtel W.;  
 RT "Complete gene map of the plastid genome of the nonphotosynthetic  
 RT euglenoid flagellate *Astasia longa*.";  
 RL Protist 151:347-351(2000).  
 CC -----  
 CC -i- SIMILARITY: BELONGS TO THE A.LONGA ORF167/ORF288 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ294725; CAC24603.1; -  
 KW Hypothetical protein; Chloroplast.  
 SQ SEQUENCE 263 AA; 32473 MW; 3508B02DB49CB4E6 CRC64;

OY 1 REVHRRP 8  
 1 :|||||  
 Db 17 RRYTHRRP 24

Query Match 71.1%; Score 32; DB 1; Length 263;  
 Best Local Similarity 50.0%; Pred. No. 22;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 7  
 YB28\_MYCTU STANDARD; PRT; 451 AA.  
 ID YB28\_MYCTU  
 AC 006580;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Hypothetical 49.3 kDa protein RV1128c.  
 GN RV1128C OR MT1160 OR MTCY2268.17C.  
 OS *Mycobacterium tuberculosis*.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; *Mycobacterium*.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
 RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and  
 RT laboratory strains.";  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 CC -i- SIMILARITY: BELONGS TO THE RV1128C/1588C/1702C/1945/3466  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z95585; CAB09039.1; -  
 DR EMBL: AE006995; AAK45417.1; -  
 DR TIGR: MT1160; -  
 DR Tuberculist; RV1128c; -  
 DR InterPro: IPR003870; DUF222.  
 DR InterPro: IPR003615; HNH\_nuc.

DR Pfam; PF02720; DUF222; 1.  
 DR SMART; SM00507; HNRc; 1.  
 KM Hypothetical protein; Complete proteome.  
 FT CONFLICT 270 270 E->G (IN REF. 2).  
 FT CONFLICT 421 451 MISSING (IN REF. 2).  
 SQ SEQUENCE 451 AA; 49274 MW; 469774E5CE148CDD CRC64;

Query Match  
 Best Local Similarity 71.1%; Score 32; DB 1; Length 451;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRFVRRP 8  
 Db 134 RRFVRRP 141

RESULT 8  
 RMS5\_MOUSE  
 ID RMS5\_MOUSE STANDARD; PRT; 551 AA.  
 AC 09EP89;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Mitochondrial 39S ribosomal protein L56 (MRP-L56) (Serine beta lactamase-like protein LACTB).  
 GN LACTB OR LACTL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV, AND BALB/C;  
 RX PubMed=11707067;  
 RA Smith T.S., Southan C., Ellington K., Campbell D., Tew D.G., RA Debouck C.;  
 RT "Identification, genomic organization, and mRNA expression of LACTB, RT encoding a serine beta-lactamase-like protein with an amino-terminal transmembrane domain.";  
 RL Genomics 78:12-14(2001).  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed predominantly in liver.  
 CC -1- SIMILARITY: BELONGS TO THE L56E FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; AF317900; AAG37910.1; -;  
 DR EMBL; AF317901; AAG37911.1; -;  
 DR MGD; MGI:1933395; Lactl.  
 KM Ribosomal protein; Mitochondrion.  
 SQ SEQUENCE 551 AA; 60705 MW; 816C1EB94B1D6009 CRC64;

Query Match  
 Best Local Similarity 71.1%; Score 32; DB 1; Length 551;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRFVRRP 8  
 Db 24 RRGARRP 31

RESULT 9  
 RPOD\_WHEAT  
 AC 09XP59; STANDARD; PRT; 1479 AA.

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE DNA-directed RNA polymerase beta" chain (EC 2.7.7.6).  
 GN RPOC2.  
 OS Triticum aestivum (wheat).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticaceae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. CHINESE SPRING;  
 RA Matsuno K., Tsunewaki K., Ohnishi Y.;  
 RT "Molecular analysis of a 21.1-kb fragment of wheat chloroplast DNA bearing RNA polymerase subunit (rpo) genes.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Chinese Spring;  
 RA Ogihara Y., Isono K., Kojima T., Endo A., Hanaoka M., Shina T., RA Terachi T., Utsugi S., Murata M., Mori N., Takumi S., Ito K., RA Gojobori T., Murali R., Murali K., Matsuno K., Ohnishi Y., Tajiri H., RA Tsunewaki K.;  
 RT "Chinese spring wheat (Triticum aestivum L.) chloroplast genome: RT complete sequence and contig clones.";  
 RL Plant Mol. Biol. Rep. 18:243-253(2000).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).  
 CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR SUBUNITS: ALPHA, BETA, BETA', AND BETA".  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; AB027572; BAB8042.1; -;  
 DR EMBL; AB042240; BAB47026.1; -;  
 DR InterPro: IPR000722; RNA\_POL\_A.  
 DR Pfam; PF00623; RNA\_POL\_A; 1.  
 KM Transferrase; Transcription; DNA-directed RNA polymerase; Chloroplast.  
 SQ SEQUENCE 1479 AA; 170043 MW; 4DFB5363CB516FB2 CRC64;

Query Match  
 Best Local Similarity 71.1%; Score 32; DB 1; Length 1479;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVRRP 8  
 Db 1429 QRFVRRP 1436

RESULT 10  
 RPOD\_MAIZE  
 ID RPOD\_MAIZE STANDARD; PRT; 1527 AA.  
 AC P16025;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE DNA-directed RNA polymerase beta" chain (EC 2.7.7.6).  
 GN RPOC2.  
 OS Zea mays (Maize).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90340289; PubMed=2381419;  
 RA Igloi G.L., Meinke A., Doery I., Koessel H.;  
 RT "Nucleotide sequence of the maize chloroplast rpo B/C1/C2 operon:  
 RT comparison between the derived protein primary structures from  
 RT various organisms with respect to functional domains.";  
 RL Mol. Gen. Genet. 221:379-394(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90175001; PubMed=2308853;  
 RA Igloi G.L., Meinke A., Doery I., Koessel H.;  
 RT "Nucleotide and derived amino acid sequence of rps2 from maize  
 RT chloroplasts.";  
 RL Nucleic Acids Res. 18:663-663(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95395841; PubMed=7666415;  
 RA Maier R.M., Neckermann K., Igloi G.L., Koessel H.;  
 RT "Complete sequence of the maize chloroplast genome: gene content,  
 RT hotspots of divergence and fine tuning of genetic information by  
 RT transcript editing.";  
 RL J. Mol. Biol. 251:614-628(1995).  
 RN [4]  
 RP SEQUENCE OF 1-52 FROM N.A.  
 RX MEDLINE=90160360; PubMed=2304916;  
 RA Hu J., Bogorad L.;  
 RT "Maize chloroplast RNA polymerase: the 180-, 120-, and 38-kilodalton  
 RT polypeptides are encoded in chloroplast genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1531-1535(1990).  
 RN [5]  
 RP SEQUENCE OF 1316-1527 FROM N.A.  
 RC STRAIN=CV. FR9CMS X RF37; TISSUE=Leaf;  
 RX MEDLINE=90272437; PubMed=2140888;  
 RA Stahl D., Rodemeier S., Subramanian A.R., Bogorad L.;  
 RT "Nucleotide sequence of a 3.46 kb region of maize chloroplast DNA  
 RT containing the gene cluster rpoC2-rps2-atpI-atpH.";  
 RL Nucleic Acids Res. 18:3073-3074(1990).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC [RNA](N).  
 CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR  
 CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X17318; CAA35197.1; -  
 DR EMBL: M31208; AA84489.1; -  
 DR EMBL: X86563; CAA60278.1; -  
 DR EMBL: X52270; CAA36511.1; -  
 DR PIR: S08248; RNZMB2.  
 DR PIR: A34846; A34846.  
 DR MaizeDB: 69586; -  
 DR Mendel: 5021; ZEMMA:rpoC2.1.  
 DR InterPro: IPR000722; RNA-pol\_A.  
 DR Pfam: PF00623; RNA-pol\_A.1.  
 DR Transferrase: Transcription; DNA-directed RNA polymerase; Chloroplast.  
 KW CONFLICT 25 25 R -> S (IN REF. 4).  
 FT ACT SITE 1527 AA; 176082 MW; 98BEA0C3165A8C3D CRC64;  
 SQ SEQUENCE

Best Local Similarity 62.5%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 RRFVRRP 8  
 Db 1466 QKRVHSP 1473  
 RESULT 11  
 ORN\_XYLFA  
 ID ORN\_XYLFA STANDARD; PRT; 193 AA.  
 AC 09PD2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Oligoribonuclease (EC 3.1.-.-).  
 GN ORN OR XF1257  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin A.C., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carreir H.,  
 RA Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kurmae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Felixto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Trufil D., Tsai S.M., Tsuchioka M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
 RL Nature 406:151-159(2000).  
 CC -1- FUNCTION: 3'-5' EXORIBONUCLEASE SPECIFIC FOR SMALL  
 CC OLIGORIBONUCLEOTIDES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE OLIGORIBONUCLEASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE003959; AAP84066.1; -  
 DR InterPro: IPR000520; Exonuclease.  
 DR Pfam: PF00929; Exonuclease; 1.  
 DR SMART: SM00479; EXOIII; 1.  
 DR Hydroxylase; Exonuclease; Nuclease; Complete proteome.  
 FT ACT SITE 135 135 POTENTIAL.  
 SQ SEQUENCE 193 AA; 21902 MW; 12D2B2C06B26D58C CRC64;

Query Match

71.1%; Score 32; DB 1; Length 1527;



Query Match 68.9%; Score 31; DB 1; Length 193;  
 Best Local Similarity 71.4%; Pred. No. 26;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
 |||:|:  
 Db 119 RRFVHRRQ 125

RESULT 12  
 VCOM\_ADECC STANDARD; PRT; 421 AA.  
 AC 065952;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE Minor core protein (Protein V).  
 GN PV.

OS Canine adenovirus type 1 (strain CLL).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=69150;

RN  
 RP SEQUENCE FROM N.A.  
 RA Campbell J.B., Zhao Y.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL: U55001; AAB05440.1; -  
 KW Core protein; Late protein.  
 SQ SEQUENCE 421 AA; 47538 MW; C84536291CDDEFP9 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 421;  
 Best Local Similarity 75.0%; Pred. No. 57;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRFVHRR 8  
 |||:|:  
 Db 67 RRFVHRRP 74

RESULT 13  
 VCOM\_ADECR STANDARD; PRT; 421 AA.  
 AC 096685;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Minor core protein (Protein V).  
 GN PV.

OS Canine adenovirus type 1 (strain RI261).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=69151;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92725900; PubMed=9129661;  
 RA Morrison M.D., Onions D.E., Nicolson L.;  
 RT "Complete DNA sequence of canine adenovirus type 1.";  
 RL J. Gen. Virol. 78:873-878(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Y07760; CAA69063.1; -  
 KW Core protein; Late protein.  
 SQ SEQUENCE 421 AA; 47649 MW; A5F523E18748211F CRC64;

Query Match 68.9%; Score 31; DB 1; Length 421;  
 Best Local Similarity 75.0%; Pred. No. 57;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRFVHRR 8  
 |||:|:  
 Db 67 RRFVHRRP 74

RESULT 14  
 YHG4\_YEAST STANDARD; PRT; 446 AA.  
 AC P38757;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DE Hypothetical 50.6 kDa protein in RPL14B-CPA1 intergenic region.  
 GN YHR004C.

OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;

RN  
 RP SEQUENCE FROM N.A.  
 RA SRRAIN=5288C; AB972;  
 RC MEDLINE=94378003; PubMed=8091229;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,  
 RA Nhan M., Rifkin L., Riles L., St. Peter H., Trevisan E., Vaughan K.,  
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
 RA Vaadin M.;  
 RT "Complete nucleotide sequence of saccharomyces cerevisiae chromosome  
 RT VIII.";  
 RL Science 265:2077-2082(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL: U10555; AAB68431.1; -  
 DR PIR: S46802; S46802.  
 DR SGD: S0001046; NEM1.  
 DR InterPro: IPR004274; NIF.  
 DR Pfam: PF03031; NIF; 1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 87 103 POTENTIAL.  
 SQ SEQUENCE 446 AA; 50641 MW; 4260748606E17062 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 446;  
 Best Local Similarity 66.7%; Pred. No. 60;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 FVHRRP 8  
 |||:|:  
 Db 295 FVHRRP 300

```

RESULT 15
PAXI_HUMAN STANDARD; PRT; 557 AA.
ID PAXI_HUMAN
AC P49023;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Paxillin.
GN PXN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-95197488; PubMed-7534286;
RA Salgia R., Li J.-L., Lo S.H., Brunkhorst B., Kansas G.S.,
RA Sobhany E.S., Sun Y., Pistick E., Hallek M., Ernst T., Tantiavahi R.,
RA Chen L.B., Griffin J.D.;
RT "Molecular cloning of human paxillin, a focal adhesion protein
RT phosphorylated by P210BCR/ABL."
RT J. Biol. Chem. 270:5039-5047(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Yamagata K., Oda N., Fututa H., Vaxillaire M., Southam L., Boriraj V.,
RA Chen X., Oda Y., Takeda J., Yamada S., Nishigori H., Lebeau M.M.,
RA Lathrop M., Cox R.D., Bell G.I.;
RT "Transcription map of the 5cm region surrounding the hepatocyte
RT nuclear factor-1a/MODY3 gene on chromosome 12."
RT Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CYTOSKELETAL PROTEIN INVOLVED IN ACTIN-MEMBRANE
CC ATTACHMENT AT SITES OF CELL ADHESION TO THE EXTRACELLULAR MATRIX
CC (FOCAL ADHESION). BINDS IN VITRO TO VINCULIN AS WELL AS TO THE SH3
CC DOMAIN OF C-SRC AND, WHEN TYROSINE PHOSPHORYLATED, TO THE SH2
CC DOMAIN OF V-CRK.
CC -!- PTM: EXTENSIVE TYROSINE PHOSPHORYLATION OCCURS DURING INTEGRIN-
CC MEDIATED CELL ADHESION, EMBRYONIC DEVELOPMENT, FIBROBLAST
CC TRANSFORMATION AND FOLLOWING STIMULATION OF CELLS BY MITOGENS.
CC -!- SIMILARITY: CONTAINS 4 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14588; AAC50104.1; -
DR EMBL; U87946; AAD00648.1; -
DR EMBL; U87941; AAD00648.1; JOINED.
DR EMBL; U87942; AAD00648.1; JOINED.
DR EMBL; U87943; AAD00648.1; JOINED.
DR EMBL; U87944; AAD00648.1; JOINED.
DR EMBL; U87945; AAD00648.1; JOINED.
DR HSSP; P04006; 1IML.
DR MIM; 602505; -
DR InterPro; IPR001781; LIM.
DR InterPro; IPR001904; Paxillin.
DR Pfam; PF00412; LIM; 4.
DR PRINTS; PR00832; PAXILLIN.
DR ProDom; PD000094; LIM; 4.
DR SMART; SM00132; LIM; 4.
DR PROSITE; PS00478; LIM DOMAIN 1; 4.
DR PROSITE; PS50023; LIM DOMAIN 2; 4.
DR Cytoskeleton; Phosphorylation; LIM domain; Repeat; Metal-binding;
KW Zinc.
FT DOMAIN 46 53 PRO-RICH.
FT MOD_RES 118 118 PHOSPHORYLATION (BY FAK) (BY SIMILARITY).
FT DOMAIN 324 374 LIM 1.
FT DOMAIN 383 433 LIM 2.
FT DOMAIN 442 492 LIM 3.

```

```

FT DOMAIN 501 551 LIM 4.
SQ SEQUENCE 557 AA; 60936 MW; B379BED638BE63A5 CRC64;

```

## Query Match

```

Best Local Similarity 57.18; Score 31; DB 1; Length 557;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 RFVHRRP 8
Db 75 RFIHQDP 81

```

```

Search completed: August 15, 2002, 11:58:17
Job time: 517 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:57:52 ; Search time 58.57 Seconds  
(without alignments)  
23.629 Million cell updates/sec

Title: US-09-613-092a-7\_COPY\_3\_10  
Perfect score: 45  
Sequence: 1 RRFVHRP 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	38	84.4	434	16 Q9RV39	Q9RV39 deinococcus
2	38	84.4	929	10 Q9LGE7	Q9LGE7 oryza sativ
3	37	82.2	127	2 Q92N73	Q92N73 streptomyc
4	35	77.8	72	9 Q64123	Q64123 bacterioph
5	35	77.8	72	16 Q31924	Q31924 bacillus su
6	35	77.8	403	16 P96402	P96402 mycobacteri
7	35	77.8	663	5 Q9VLG3	Q9VLG3 drosophila
8	34	75.6	80	16 Q9JYH1	Q9JYH1 neisseria m
9	34	75.6	183	2 Q93P28	Q93P28 pseudomonas
10	34	75.6	217	5 Q95TP9	Q95TP9 drosophila
11	34	75.6	243	5 Q9VRG7	Q9VRG7 drosophila
12	34	75.6	244	4 Q95476	Q95476 homo sapien
13	34	75.6	244	4 Q96G09	Q96G09 homo sapien
14	34	75.6	254	5 Q95XU1	Q95XU1 caenorhabdi
15	34	75.6	322	2 Q9F2I8	Q9F2I8 synechococc
16	34	75.6	386	10 Q94G10	Q94G10 oryza sativ

17	33	73.3	118	2 Q9S2K1	Q9S2K1 streptomyc
18	33	73.3	241	10 Q9LPB6	Q9LPB6 arabidopsis
19	33	73.3	246	10 Q9M234	Q9M234 arabidopsis
20	33	73.3	320	5 Q9N867	Q9N867 leishmania
21	33	73.3	548	12 Q9YMW2	Q9YMW2 lymantria d
22	33	73.3	567	16 Q9ZM37	Q9ZM37 rhizobium m
23	33	73.3	577	16 Q9Z9B5	Q9Z9B5 chlamydia p
24	33	73.3	702	2 Q9AD11	Q9AD11 streptomyc
25	33	73.3	722	5 Q61208	Q61208 caenorhabdi
26	33	73.3	786	5 Q9P653	Q9P653 neurospora
27	32	71.1	57	2 Q86472	Q86472 rhodobacter
28	32	71.1	60	2 Q83021	Q83021 rhodobacter
29	32	71.1	138	17 Q9YBN3	Q9YBN3 aeropyrum p
30	32	71.1	145	11 Q9DA47	Q9DA47 homo musculi
31	32	71.1	174	4 Q9P0B9	Q9P0B9 homo sapien
32	32	71.1	202	10 Q80513	Q80513 arabidopsis
33	32	71.1	214	16 Q98J73	Q98J73 rhizobium l
34	32	71.1	295	16 Q53514	Q53514 mycobacteri
35	32	71.1	301	5 Q9VYF8	Q9VYF8 drosophila
36	32	71.1	321	10 Q9LVR5	Q9LVR5 arabidopsis
37	32	71.1	321	10 Q94B21	Q94B21 arabidopsis
38	32	71.1	334	2 Q9AK99	Q9AK99 streptomyc
39	32	71.1	346	4 Q96DV4	Q96DV4 homo sapien
40	32	71.1	369	16 Q92NFO	Q92NFO rhizobium m
41	32	71.1	370	10 Q9S7D5	Q9S7D5 arabidopsis
42	32	71.1	443	2 Q93S73	Q93S73 myxococcus
43	32	71.1	460	10 Q9AWS1	Q9AWS1 oryza sativ
44	32	71.1	523	10 Q9ANQ9	Q9ANQ9 oryza sativ
45	32	71.1	527	2 Q9S201	Q9S201 streptomyc

ALIGNMENTS

RESULT 1  
Q9RV39 PRELIMINARY; PRT; 434 AA.  
ID Q9RV39  
AC Q9RV39;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE tRNA NUCLEOTIDYLTRANSFERASE/POLY A POLYMERASE FAMILY PROTEIN.  
GN DR1191.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP  
RC STRAIN=RL;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Morfitt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Ueberback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RA "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans RL.";  
RL Science 286:1571-1577(1999).  
DR EMBL: AE001968; AAF10763.1; -  
DR TIGR: DR1191; -  
DR InterPro: IPR002819; HD.  
DR InterPro: IPR003607; HDC.  
DR InterPro: IPR002646; POLYA\_pol.  
DR Pfam: PF01966; HD; 1.  
DR Pfam: PF01743; POLYA\_pol; 1.  
DR SMART: SM00471; HDC; 1.  
KW Transferase; Complete proteome.  
SQ SEQUENCE 434 AA: 47987 MW; A792A1A6DD6A0D2 CRC64;

Query Match 84.4%; Score 38; DB 16; Length 434;

Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRFVHRR 7  
1111111  
Db 322 RRFVHRR 328

## RESULT 2

09LGE7 PRELIMINARY; PRT; 929 AA.  
AC 09LGE7;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2001 (TReMBLrel. 15, Last sequence update)  
DE P0406H10.1 PROTEIN (P0509B06.3 PROTEIN).  
GN P0406H10.1 OR P0509B06.3.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
clone: P0406H10.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
clone: P0509B06.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP002824; BAB07977.1; -;  
DR EMBL: AP002803; BAB63468.1; -;  
DR InterPro: IPR002794; DUF92.  
DR Pfam: PF01940; DUF92; 1.  
DR ProDom: PD014594; DUF92; 1.  
SQ SEQUENCE 929 AA; 100026 MW; FE869AD16F3F6A78 CRC64;

Query Match 84.4%; Score 38; DB 10; Length 929;  
Best Local Similarity 87.5%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRFVHRR 8  
1111111  
Db 901 RRFVHRR 908

## RESULT 3

092N73 PRELIMINARY; PRT; 127 AA.  
AC 092N73;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DE 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
DE MEIC1.  
GN MEIC1.  
OS Streptomyces griseus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IFO13350;  
RA Ueda K., Beppu T.;  
RT "Tyrosinase gene from Streptomyces griseus.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB022095; BAA37084.1; -;

SQ SEQUENCE 127 AA; 13707 MW; 93801073CCAC86CA CRC64;

Query Match 82.2%; Score 37; DB 2; Length 127;  
Best Local Similarity 87.5%; Pred. No. 7.2;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRFVHRR 8  
1111111  
Db 66 RRFVHRR 73

## RESULT 4

064123 PRELIMINARY; PRT; 72 AA.  
AC 064123;  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 8.4 KDA PROTEIN.  
GN YOON.  
OS Bacteriophage SPBc2.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=66797;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.,  
RA Karanata D.;  
RT "The complete nucleotide sequence of the Bacillus subtilis spBc2  
prophage.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF020713; AAC13083.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 72 AA; 8382 MW; 11178AF395985ED8 CRC64;

Query Match 77.8%; Score 35; DB 9; Length 72;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FVHRRP 8  
1111111  
Db 60 FVHRRP 65

## RESULT 5

031924 PRELIMINARY; PRT; 72 AA.  
AC 031924;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DE 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE YOON PROTEIN.  
GN YOON.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriss R., Boursier L., Brans A., Braun M., Bridgnell S.C., Bron S.,  
RA Brouillet S., Buschi C.V., Caldwell B., Capiano V., Carter N.M.,  
RA Choi S.K., Codani J.U., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klairr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nohack M.,  
RA Moore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M.,  
RA Paresan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadleir Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Seliguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,  
RA Viari A., Wambler R., Wedler E., Wedler H., Weitenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,  
RT "The complete genome sequence of the gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: 299114; CAB13949.1; -  
KM Complete proteome.  
SQ SEQUENCE 72 AA: 8382 MW: 11178AF395985ED8 CRC64;

Query Match 77.8%; Score 35; DB 16; Length 72;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 FVHRP 8  
111111  
Db 60 FVHRP 65

RESULT 6  
P96402 PRELIMINARY; PRT; 403 AA.  
AC P96402;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE HYDROTHERMAL 44.3 KDA PROTEIN.  
GN LIPC OR RV0220 OR MTCY08D5.15.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Horsley T., Jagels K., Kiroh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLSTERASE/LIPASE FAMILY.  
DR EMBL: 292669; CAB07015.1; -  
DR Tuberculist; RV0220; -  
DR InterPro; IPR0002018; Carboxylesterase\_B.  
DR InterPro; IPR000379; Est\_lip\_chioest\_cactsite.  
DR PROSITE; PS00122; CARBOXYLSTERASE\_B\_1; 1.

KW Complete proteome; Hydrolase; Hypothetical protein.  
SQ SEQUENCE 403 AA: 44307 MW: 4FBAB5F1DECD563 CRC64;

Query Match 77.8%; Score 35; DB 16; Length 403;  
Best Local Similarity 85.7%; Pred. No. 53;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RRVHRR 7  
111111  
Db 125 RRVHRR 131

RESULT 7  
O9VLG3 PRELIMINARY; PRT; 663 AA.  
AC O9VLG3;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE CG18088 PROTEIN.  
GN CG18088.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,  
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003623; AAF52729.1; -  
DR FlyBase; FBgn0032082; CG18088.  
DR InterPro; IPR004245; DUF229.  
DR Pfam; PF02995; DUF229; 1.

SO SEQUENCE ~ 663 AA; 77363 MW; 135B5E43A61591B2 CRC64;

Query Match  
Best Local Similarity 77.8%; Score 35; DB 5; Length 663;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVHRR 8  
Db 369 KRFVHRR 376

RESULT 8

09JYH1 PRELIMINARY; PRT; 80 AA.  
AC 09JYH1;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE HYPOTHETICAL PROTEIN NMB1589.  
GN NMB1589.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MC58 / SEROGROUP B;  
RA MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Clecco A., Parksey D.S., Blair E., Ciftore H., Clark E.B.,  
RA Cotton M.D., Utechtack T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
MC58.";  
RT NC58."  
RL Science 287:1809-1815(2000).  
DR EMBL; AEO02509; AAF41942.1; -.  
DR TIGR; NMB1589; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 80 AA; 9395 MW; 8FBC5E79AB29DCEA CRC64;

Query Match  
Best Local Similarity 75.6%; Score 34; DB 16; Length 80;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
Db 63 KRFVHRR 69

RESULT 9

093PZ8 PRELIMINARY; PRT; 183 AA.  
AC 093PZ8;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE RSP.  
GN RSP.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SBW25;  
RA "Type III secretion in plant growth-promoting Pseudomonas fluorescens  
SBW25.";  
RT

RL MOL. Microbiol. 0:0-0(2001).  
DR EMBL; AF292566; AAK81917.1; -.  
SQ SEQUENCE 183 AA; 21443 MW; 47589EA31FBE866 CRC64;

Query Match  
Best Local Similarity 75.6%; Score 34; DB 2; Length 183;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 RRFVHRR 7  
Db 31 RRFVHRR 37

RESULT 10

095TP9 PRELIMINARY; PRT; 217 AA.  
AC 095TP9;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE ID31286P.  
GN RPS6.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Y, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Fartan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuncio J., Paclob J., Paragas V., Park S., Phuanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY058620; AAL13849.1; -.  
SQ SEQUENCE 217 AA; 24696 MW; C2081AE009A2DD8E CRC64;

Query Match  
Best Local Similarity 75.6%; Score 34; DB 5; Length 217;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVHRR 8  
Db 123 RRFVHRR 130

RESULT 11

09VRG7 PRELIMINARY; PRT; 243 AA.  
AC 09VRG7;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE CG1696 PROTEIN.  
GN CG1696.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,



RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Adili J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Beron P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gotrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hariri N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrstkas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AF003568; AAF50833.1;  
 DR FlyBase: FBgn0031173; CG1696.  
 DR InterPro: IPR004274; NIF.  
 DR Pfam: PF03031; NIF; 1.  
 DR PIRam: IPR004274; NIF.  
 SQ SEQUENCE 243 AA; 28486 MW; 1E278DD1D8DF60C6 CRC64;

Query Match 75.6%; Score 34; DB 5; Length 243;  
 Best Local Similarity 75.0%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RRFVHRRP 8  
 Db 104 RRFVHRRP 111  
 RESULT 12  
 ID 095476 PRELIMINARY; PRT; 244 AA.  
 AC 095476;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHEICAL 28.3 KDA PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA Keen J., Inglehearn C.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ011916; CAA09865.1;  
 DR InterPro: IPR004274; NIF.  
 DR Pfam: PF03031; NIF; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 244 AA; 28347 MW; E23843B815DEFDF0 CRC64;

Query Match 75.6%; Score 34; DB 4; Length 244;  
 Best Local Similarity 75.0%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RRFVHRRP 8  
 Db 105 RRFVHRRP 112  
 RESULT 13  
 ID 096G09 PRELIMINARY; PRT; 244 AA.  
 AC 096G09;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE UNKNOWN (PROTEIN FOR MCC:16648).  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;  
 RA Strausberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC009295; AAH09295.1;  
 SQ SEQUENCE 244 AA; 28377 MW; 062952A90F74575A CRC64;

Query Match 75.6%; Score 34; DB 4; Length 244;  
 Best Local Similarity 75.0%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RRFVHRRP 8  
 Db 105 RRFVHRRP 112  
 RESULT 14  
 ID 095XU1 PRELIMINARY; PRT; 254 AA.  
 AC 095XU1;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHEICAL 27.6 KDA PROTEIN.  
 GN Y67DBA.3.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for  
 RT investigating biology. The *C. elegans* Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Edwards J., Lamar B., Minx P., Du H., Kemp K., Wohlmann P.,  
 RA Walker C.;  
 RT "The sequence of *C. elegans* cosmid Y67DBA.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC024848; AAK68545.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 254 AA; 27600 MW; 32895C8BDF425785 CRC64;

Query Match 75.6%; Score 34; DB 5; Length 254;  
 Best Local Similarity 73.0%; Pred. No. 53;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRFVHRP 8  
 ||| ||:|  
 DB 73 RRFVHRP 80

## RESULT 15

O9F218 PRELIMINARY; PRT; 322 AA.  
 ID O9F218  
 AC O9F218;  
 DT 01-MAR-2001 (TRENBLREL. 16, Created)  
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE CYANOPHYCINASE.  
 GN CPHB.  
 OS Synecoccus elongatus (Thermosynechococcus elongatus).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=32046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20408669; PubMed-10951215;  
 RA Berg H., Ziegler K., Piotukh K., Baier K., Lockau W.,  
 RA Volkermer-Engert R.;  
 RT "Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-  
 RT poly-L-aspartic acid (cyanophycin): Mechanism of the cyanophycin  
 RT synthetase reaction studied with synthetic primers.";  
 RL Eur. J. Biochem. 267:5561-5570(2000).  
 DR EMBL: AJ288949; CAC07986.1; -.  
 SQ SEQUENCE 322 AA; 35502 MW; 8D319403DA5245B7 CRC64;

Query Match 75.6%; Score 34; DB 2; Length 322;  
 Best Local Similarity 71.4%; Pred. No. 67;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RRFVHRP 8  
 ||:|| |  
 DB 28 RRFVHRP 34

Search completed: August 15, 2002, 11:57:54  
 Job time: 529 sec

1	91	100.0	15	20	AAV30351	Epitope derived fr
2	47	51.6	73	22	ABG05173	Novel human diagno
3	46	50.5	113	22	AAU52411	Proionibacterium
4	45	49.5	84	22	AAG77743	Human colon cancer
5	45	49.5	185	21	AAI14880	Aradidopsis thallia
6	45	49.5	436	21	AAAG30975	Aradidopsis thallia
7	45	49.5	445	21	AAAG2039	Aradidopsis thallia
8	45	49.5	454	21	AAAG30974	Aradidopsis thallia
9	45	49.5	463	21	AAAG2038	Aradidopsis thallia
10	44	48.4	21	22	ABG28607	Novel human diagno
11	44	48.4	395	22	ABG04577	Novel human diagno

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Ades EM, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;  
PI Zeller JL.  
XX  
XX WPI: 1999-540849/45.  
XX  
XX  
XX New peptides corresponding to Streptococcus pneumoniae Psaa, used  
PT for treating or preventing Streptococcus pneumoniae infection in a  
PT subject  
XX

PS Claim 6; Page 43; 58pp; English.  
XX  
CC AAY30351-54 represent immunogenic peptides which are derived from  
CC a pneumococcal surface adhesion A protein (PsaA). The specification  
CC describes monoclonal antibodies which bind epitopes of the PsaA protein  
CC (e.g. present sequence). The peptides can be used in vaccines to prevent  
CC Streptococcus pneumoniae infections. The antibodies of the invention  
CC can also be used to detect S. pneumoniae in a sample or individual.  
XX  
SQ Sequence 15 AA;  
  
Query Match 100.0%; Score 91; DB 20; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.8e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TVSRVPTAWAFHGY 15  
Db 1 tvsrvpwtafhgy 15  
  
RESULT 2  
ID ABG05173 standard; Protein: 73 AA.  
XX  
AC ABG05173;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #5164.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS69360.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 35532; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG03077 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 73 AA;  
  
Query Match 51.6%; Score 47; DB 22; Length 73;  
Best Local Similarity 50.0%; Pred. No. 6.3;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
OY 2 VSRVPWTAWAFHGY 15  
Db 60 vskfpwawpfsry 73  
  
RESULT 3  
ID AAU52411 standard; Protein: 113 AA.  
XX  
AC AAU52411;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #13307.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteophthic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN W0200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI; 2001-616774/71.  
DR N-PSDB; AAS59554.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID No 13606; 106pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies

CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 113 AA;

Query Match 50.5%; Score 46; DB 22; Length 113;  
Best Local Similarity 58.3%; Pred. No. 14;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TVSRVPTWMAF 12  
| | : | | | : |  
Db 27 tcsmpwtpmaw 38

## RESULT 4

AAAG77743  
ID AAAG77743 standard; Protein; 84 AA.

AC AAAG77743;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO: 8509.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26524.

PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI; 2001-235357/24.

DR N-PSDB; AAH37150.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
XX Claim 11; Page 9762-9763; 9803pp; English.

XX AAH32943 to AAH37195 and AAAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate P  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing  
XX inactive proteins or to supplement the patient's own production of P.  
XX Additionally, N may be used to produce the colon cancer-associated Ps,  
XX by inserting the nucleic acids into a host cell and culturing the cell  
XX to express the proteins. N and P can be used in the prevention, diagnosis  
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
XX and AAAG77789 represent sequences used in the exemplification of the  
XX present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX  
SQ Sequence 84 AA;

Query Match 49.5%; Score 45; DB 22; Length 84;  
Best Local Similarity 75.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 PWTMAFH 13  
| | | | | |  
Db 63 pwtawxxh 70

## RESULT 5

AAAG14880  
ID AAAG14880 standard; Protein; 185 AA.

AC AAAG14880;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 14909.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132048.

PR 06-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0136021.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR	01-JUN-1999;	990S-0137222.
PR	03-JUN-1999;	990S-0137528.
PR	04-JUN-1999;	990S-0137502.
PR	07-JUN-1999;	990S-0137724.
PR	08-JUN-1999;	990S-0138094.
PR	10-JUN-1999;	990S-0138540.
PR	10-JUN-1999;	990S-0138847.
PR	14-JUN-1999;	990S-0139119.
PR	16-JUN-1999;	990S-0139453.
PR	16-JUN-1999;	990S-0139452.
PR	17-JUN-1999;	990S-0139454.
PR	18-JUN-1999;	990S-0139455.
PR	18-JUN-1999;	990S-0139456.
PR	18-JUN-1999;	990S-0139457.
PR	18-JUN-1999;	990S-0139458.
PR	18-JUN-1999;	990S-0139460.
PR	18-JUN-1999;	990S-0139461.
PR	18-JUN-1999;	990S-0139462.
PR	18-JUN-1999;	990S-0139463.
PR	18-JUN-1999;	990S-0139750.
PR	21-JUN-1999;	990S-0139763.
PR	22-JUN-1999;	990S-0139817.
PR	22-JUN-1999;	990S-0139899.
PR	23-JUN-1999;	990S-0140353.
PR	23-JUN-1999;	990S-0140354.
PR	24-JUN-1999;	990S-0140695.
PR	28-JUN-1999;	990S-0140823.
PR	29-JUN-1999;	990S-0140991.
PR	30-JUN-1999;	990S-0141287.
PR	01-JUL-1999;	990S-0141842.
PR	01-JUL-1999;	990S-0142154.
PR	02-JUL-1999;	990S-0142055.
PR	06-JUL-1999;	990S-0142390.
PR	08-JUL-1999;	990S-0142803.
PR	09-JUL-1999;	990S-0142920.
PR	12-JUL-1999;	990S-0142977.
PR	13-JUL-1999;	990S-0143542.
PR	14-JUL-1999;	990S-0143624.
PR	15-JUL-1999;	990S-0144005.
PR	16-JUL-1999;	990S-0144085.
PR	16-JUL-1999;	990S-0144086.
PR	19-JUL-1999;	990S-0144325.
PR	19-JUL-1999;	990S-0144351.
PR	19-JUL-1999;	990S-0144332.
PR	19-JUL-1999;	990S-0144333.
PR	19-JUL-1999;	990S-0144334.
PR	20-JUL-1999;	990S-0144335.
PR	20-JUL-1999;	990S-0144352.
PR	20-JUL-1999;	990S-0144632.
PR	20-JUL-1999;	990S-0144884.
PR	21-JUL-1999;	990S-0144884.
PR	21-JUL-1999;	990S-0145086.
PR	21-JUL-1999;	990S-0145088.
PR	22-JUL-1999;	990S-0145085.
PR	22-JUL-1999;	990S-0145087.
PR	22-JUL-1999;	990S-0145089.
PR	23-JUL-1999;	990S-0145192.
PR	23-JUL-1999;	990S-0145145.
PR	23-JUL-1999;	990S-0145218.
PR	23-JUL-1999;	990S-0145224.
PR	26-JUL-1999;	990S-0145276.
PR	27-JUL-1999;	990S-0145913.
PR	27-JUL-1999;	990S-0145918.
PR	27-JUL-1999;	990S-0145919.
PR	28-JUL-1999;	990S-0145951.
PR	02-AUG-1999;	990S-0146386.
PR	02-AUG-1999;	990S-0146388.
PR	02-AUG-1999;	990S-0146389.
PR	03-AUG-1999;	990S-0147038.
PR	04-AUG-1999;	990S-0147204.
PR	04-AUG-1999;	990S-0147302.
<hr/>		
PR	05-AUG-1999;	990S-0147192.
PR	05-AUG-1999;	990S-0147260.
PR	06-AUG-1999;	990S-0147303.
PR	06-AUG-1999;	990S-0147416.
PR	09-AUG-1999;	990S-0147493.
PR	09-AUG-1999;	990S-0147935.
PR	10-AUG-1999;	990S-0148171.
PR	11-AUG-1999;	990S-0148319.
PR	12-AUG-1999;	990S-0148341.
PR	13-AUG-1999;	990S-0148565.
PR	13-AUG-1999;	990S-0148368.
PR	16-AUG-1999;	990S-0149175.
PR	17-AUG-1999;	990S-0149426.
PR	18-AUG-1999;	990S-0149426.
PR	20-AUG-1999;	990S-0149722.
PR	20-AUG-1999;	990S-0149723.
PR	20-AUG-1999;	990S-0149929.
PR	23-AUG-1999;	990S-0149902.
PR	23-AUG-1999;	990S-0149930.
PR	23-AUG-1999;	990S-0150566.
PR	26-AUG-1999;	990S-0150884.
PR	26-AUG-1999;	990S-0151065.
PR	27-AUG-1999;	990S-0151066.
PR	27-AUG-1999;	990S-0151080.
PR	30-AUG-1999;	990S-0151303.
PR	31-AUG-1999;	990S-0151438.
PR	01-SEP-1999;	990S-0151930.
PR	07-SEP-1999;	990S-0152363.
PR	10-SEP-1999;	990S-0153070.
PR	13-SEP-1999;	990S-0153758.
PR	15-SEP-1999;	990S-0154018.
PR	16-SEP-1999;	990S-0154039.
PR	20-SEP-1999;	990S-0154779.
PR	22-SEP-1999;	990S-0155139.
PR	23-SEP-1999;	990S-0155486.
PR	24-SEP-1999;	990S-0155659.
PR	28-SEP-1999;	990S-0156458.
PR	29-SEP-1999;	990S-0156556.
PR	04-OCT-1999;	990S-0157117.
PR	05-OCT-1999;	990S-0157753.
PR	06-OCT-1999;	990S-0157865.
PR	07-OCT-1999;	990S-0158029.
PR	08-OCT-1999;	990S-0158232.
PR	12-OCT-1999;	990S-0158359.
PR	13-OCT-1999;	990S-0159293.
PR	13-OCT-1999;	990S-0159294.
PR	13-OCT-1999;	990S-0159295.
PR	14-OCT-1999;	990S-0159329.
PR	14-OCT-1999;	990S-0159330.
PR	14-OCT-1999;	990S-0159331.
PR	14-OCT-1999;	990S-0159637.
PR	14-OCT-1999;	990S-0159638.
PR	18-OCT-1999;	990S-0159584.
PR	21-OCT-1999;	990S-0160741.
PR	21-OCT-1999;	990S-0160767.
PR	21-OCT-1999;	990S-0160768.
PR	21-OCT-1999;	990S-0160770.
PR	21-OCT-1999;	990S-0160814.
PR	22-OCT-1999;	990S-0160815.
PR	22-OCT-1999;	990S-0160980.
PR	22-OCT-1999;	990S-0160981.
PR	22-OCT-1999;	990S-0160989.
PR	25-OCT-1999;	990S-0161404.
PR	25-OCT-1999;	990S-0161405.
PR	25-OCT-1999;	990S-0161406.
PR	26-OCT-1999;	990S-0161359.
PR	26-OCT-1999;	990S-0161360.
PR	26-OCT-1999;	990S-0161361.
PR	28-OCT-1999;	990S-0161920.
PR	28-OCT-1999;	990S-0161992.
PR	28-OCT-1999;	990S-0161993.
PR	29-OCT-1999;	990S-0162142.

Query Match 49.5%; Score 45; DB 21; Length 185;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 PWTAAAF 12  
| | | | | |  
DB 32 pwtaway 38

RESULT 6  
AAG30975  
ID AAG30975 standard; Protein; 436 AA.  
XX  
AC AAG30975;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37124.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158359.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

49.5%; Score 45; DB 21; Length 436;  
Best Local Similarity 85.7%; Pred. No. 79;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 6 PWTAMAF 12  
| | | | |

Db 14 pwtaway 20

## RESULT 7

AAG52039  
ID AAG52039 standard; Protein: 445 AA.

XX AC AAG52039;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 66108.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.





XX	17-OCT-2000	(first entry)	PR	18-JUN-1999;	99US-0139750.
DT			PR	18-JUN-1999;	99US-0139763.
XX		Arabidopsis thaliana protein fragment SEQ ID NO: 37123.	PR	21-JUN-1999;	99US-0139817.
DE			PR	22-JUN-1999;	99US-0139839.
XX		Protein identification: signal transduction pathway; metabolic pathway;	PR	23-JUN-1999;	99US-0140353.
KW		hybridisation assay; genetic mapping; gene expression control; promoter;	PR	23-JUN-1999;	99US-0140354.
KW		termination sequence.	PR	24-JUN-1999;	99US-0140699.
XX		Arabidopsis thaliana.	PR	28-JUN-1999;	99US-0140823.
OS			PR	29-JUN-1999;	99US-0140991.
XX	EPI033405-A2.		PR	30-JUN-1999;	99US-0141287.
PN			PR	01-JUL-1999;	99US-0141842.
XX			PR	01-JUL-1999;	99US-0142154.
PD	06-SEP-2000.		PR	02-JUL-1999;	99US-0142055.
XX			PR	06-JUL-1999;	99US-0142390.
XX			PR	08-JUL-1999;	99US-0142803.
PE	25-FEB-2000; 2000EP-0301439.		PR	09-JUL-1999;	99US-0142920.
XX			PR	12-JUL-1999;	99US-0142977.
XX			PR	13-JUL-1999;	99US-0143542.
XX			PR	14-JUL-1999;	99US-0143624.
PR	25-FEB-1999;	99US-0121825.	PR	15-JUL-1999;	99US-0144005.
PR	05-MAR-1999;	99US-0123180.	PR	16-JUL-1999;	99US-0144085.
PR	09-MAR-1999;	99US-0123548.	PR	16-JUL-1999;	99US-0144086.
PR	23-MAR-1999;	99US-0125788.	PR	19-JUL-1999;	99US-0144325.
PR	25-MAR-1999;	99US-0126264.	PR	19-JUL-1999;	99US-0144331.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144332.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144333.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144334.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144335.
PR	16-APR-1999;	99US-0129845.	PR	20-JUL-1999;	99US-0144352.
PR	19-APR-1999;	99US-0130077.	PR	20-JUL-1999;	99US-0144632.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144864.
PR	23-APR-1999;	99US-0130510.	PR	21-JUL-1999;	99US-0144814.
PR	23-APR-1999;	99US-0130891.	PR	21-JUL-1999;	99US-0145086.
PR	28-APR-1999;	99US-0131449.	PR	22-JUL-1999;	99US-0145085.
PR	30-APR-1999;	99US-0132048.	PR	22-JUL-1999;	99US-0145087.
PR	30-APR-1999;	99US-0132407.	PR	22-JUL-1999;	99US-0145089.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145192.
PR	05-MAY-1999;	99US-0132485.	PR	23-JUL-1999;	99US-0145145.
PR	06-MAY-1999;	99US-0132486.	PR	23-JUL-1999;	99US-0145224.
PR	07-MAY-1999;	99US-0132487.	PR	23-JUL-1999;	99US-0145226.
PR	11-MAY-1999;	99US-0134256.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134218.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145918.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145919.
PR	18-MAY-1999;	99US-0134768.	PR	28-JUL-1999;	99US-0145951.
PR	19-MAY-1999;	99US-0134941.	PR	02-AUG-1999;	99US-0146386.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146388.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146389.
PR	24-MAY-1999;	99US-0135629.	PR	03-AUG-1999;	99US-0147038.
PR	25-MAY-1999;	99US-0136021.	PR	04-AUG-1999;	99US-0147204.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147302.
PR	28-MAY-1999;	99US-0136782.	PR	05-AUG-1999;	99US-0147192.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147260.
PR	03-JUN-1999;	99US-0137528.	PR	06-AUG-1999;	99US-0147303.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147416.
PR	07-JUN-1999;	99US-0137724.	PR	09-AUG-1999;	99US-0147493.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138540.	PR	10-AUG-1999;	99US-0148319.
PR	10-JUN-1999;	99US-0138847.	PR	11-AUG-1999;	99US-0148319.
PR	14-JUN-1999;	99US-0139119.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139452.	PR	13-AUG-1999;	99US-0148565.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148684.
PR	17-JUN-1999;	99US-0139492.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139454.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139461.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139462.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139463.			

PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151330.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 49.5%; Score 45; DB 21; Length 454;  
Best Local Similarity 85.7%; Pred. No. 82;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 PWTMAF 12  
Db 32 pwtway 38

RESULT 9  
AAGS2038  
ID AAGS2038 standard; Protein; 463 AA.

XX AAGS2038;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 66107.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.  
XX Arabidopsis thaliana.  
OS EP1033405-A2.  
XX 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138640.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145195.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149388.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 16-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 49.5%; Score 45; DB 21; Length 463;  
Best Local Similarity 85.7%; Pred. No. 84;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PWTAMAF 12  
|||||:  
Db 32 pwtaway 38

RESULT 10  
ABG28607  
ID ABG28607 standard; Protein; 21 AA.  
XX  
AC ABG28607;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #28598.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
OS  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX

```

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB: AAS92794.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 58966; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 21 AA:
SQ

```

```

Query Match 48.4%; Score 44; DB 22; Length 21;
Best Local Similarity 35.7%; Pred. No. 4.8;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

OY 2 VSRVPTWTAAPHGY 15
   : : : : : : : :
Db 8 lylklpwavvsfsqy 21

```

```

RESULT 11
ABG04577
ID ABG04577 standard; Protein; 395 AA.
XX
XX ABG04577;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #4568.
DE
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF

```

```

XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB: AAS68764.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 34936; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 395 AA:
SQ

```

```

Query Match 48.4%; Score 44; DB 22; Length 395;
Best Local Similarity 66.7%; Pred. No. 16+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 6 PWTWMAFHG 14
   : : : : : : : :
Db 386 pwtwmafhg 394

```

```

RESULT 12
AAR37442
ID AAR37442 standard; Protein; 602 AA.
XX
XX AAR37442;
AC
XX
XX 06-OCT-1993 (first entry)
DT
XX
XX Full-length human pseudocholinesterase.
DE
XX
XX butyrlcholinesterase; acylcholine acylhydrolase; EC3.1.1.8; psi-CHE;
KW pseudo-CHE; neurotransmitter; organophosphorus insecticide; OP-poison;
KW antidote.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..24
FT /note= "putative leader peptide"
FT Modified-site 45..47
FT /note= "potential N-glycosylation site"

```

```

FT Modified-site 134..136
FT /note= "potential N-glycosylation site"
FT Modified-site 269..271
FT /note= "potential N-glycosylation site"
FT Modified-site 284..286
FT /note= "potential N-glycosylation site"
FT Modified-site 369..371
FT /note= "potential N-glycosylation site"
FT Modified-site 509..511
FT /note= "potential N-glycosylation site"
FT Modified-site 514..516
FT /note= "potential N-glycosylation site"
FT Active-site 226
FT /note= "active site Serine"
XX
XX US5215909-A.
XX
XX PD 01-JUN-1993.
XX
XX PE 18-JUN-1986; 86US-0875737.
XX
XX PR 18-JUN-1986; 86US-0875737.
XX 21-AUG-1987; 87US-0087724.
XX 15-AUG-1990; 90US-0572911.
XX
XX PA (YEDA ) YEDA RES & DEV CO LTD.
XX
XX PI Score H;
XX
XX DR WPI: 1993-188509/23.
XX DR N-PSDB; AAQ42496.
XX
XX PT Recombinant human gene encoding human pseudo-cholinesterase -
XX used to treat organo-phosphorus poisoning
XX
XX PS Disclosure; Columns 35-40; 34pp; English.
XX
XX CC A cDNA library prepared from foetal brain mRNA was screened with
XX degenerate probe pools based on the organophosphorus binding site of
XX cholinesterases. A 764 nucleotide insert (designated FBCH12) was
XX isolated from one positive clone and sequenced. This insert (AAQ42495),
XX containing an ORF large enough to code for about half the subunit
XX size of human cholinesterase, was used as a probe to obtain the full-
XX length pseudocholinesterase sequence (AAQ42496).
XX
XX SQ Sequence 602 AA;

```

Query Match 48.4%; Score 44; DB 14; Length 602;  
Best Local Similarity 46.7%; Pred. No. 1.5e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

```

QY 3 SRVPWTAW-AFHGX 15
   1::||| | |||
Db 454 sklpwemwymghy 468

```

RESULT 13  
ID AAY44573  
XX AAY44573 standard; Protein; 602 AA.  
XX  
XX AC AAY44573;  
XX  
XX DT 04-APR-2000 (first entry)  
XX  
XX DE Human wild type Butyrylcholinesterase (BCHE) protein.  
XX  
XX KW Butyrylcholinesterase; BCHE allele; neurological disease; treatment;  
XX therapy; allelic variant; BCHE-K; apoE4 allele; neurofibromatosis;  
XX non-AD neurological disease; Alzheimer's disease; Huntington's disease;  
XX depression; amyotrophic lateral sclerosis; multiple sclerosis; stroke;  
XX Parkinson's disease; multi-infarct dementia; human.

```

OS Homo sapiens.
XX
XX PN WO9966072-A2.
XX
XX PD 23-DEC-1999.
XX
XX PE 16-JUN-1999; 99WO-IB01298.
XX
XX PR 16-JUN-1998; 98US-0089406.
XX
XX PA (NOVA-) NOVA MOLECULAR INC.
XX
XX PI Seviigny P, Wiedusch H, Schappert K;
XX WPI: 2000-126550/11.
XX DR N-PSDB; AAZ49470.
XX
XX PT Prediction of drug efficacy for treating neurological diseases like
XX Alzheimer's disease, neurofibromatosis, Huntington's disease -
XX
XX PS Example 1; Fig 3; 37pp; English.
XX
XX CC The present sequence is the wild type human butyrylcholinesterase (BCHE)
XX protein. Determining BCHE allele status of a patient helps predicting
XX risk for neurological diseases' efficacy of therapy and determining
XX treatment protocol. Presence of BCHE allelic variant, BCHE-K and
XX apoE4 allele indicate patient's risk for having a neurological
XX disease. This method enables treating Alzheimer's disease, depression,
XX neurofibromatosis, Huntington's disease, amyotrophic lateral sclerosis,
XX multiple sclerosis, stroke, Parkinson's disease, multi-infarct dementia
XX and other non-AD neurological diseases.
XX
XX SQ Sequence 602 AA;

```

Query Match 48.4%; Score 44; DB 21; Length 602;  
Best Local Similarity 46.7%; Pred. No. 1.5e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

```

QY 3 SRVPWTAW-AFHGX 15
   1::||| | |||
Db 454 sklpwemwymghy 468

```

RESULT 14  
ID AAY44574  
XX AAY44574 standard; Protein; 602 AA.  
XX  
XX AC AAY44574;  
XX  
XX DT 04-APR-2000 (first entry)  
XX  
XX DE Human Butyrylcholinesterase-K (BCHE-K) protein.  
XX  
XX KW Butyrylcholinesterase-K; BCHE-K; BCHE allele; neurological disease;  
XX therapy; treatment; allelic variant; apoE4 allele; neurofibromatosis;  
XX non-AD neurological disease; Alzheimer's disease; Huntington's disease;  
XX depression; amyotrophic lateral sclerosis; multiple sclerosis; stroke;  
XX Parkinson's disease; multi-infarct dementia; human.  
XX  
XX OS Homo sapiens.  
XX  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 567 /note= "Wild type Ala replaced with Thr"  
XX  
XX PN WO9966072-A2.  
XX  
XX PD 23-DEC-1999.  
XX  
XX PF 16-JUN-1999; 99WO-IB01298.  
XX  
XX PR 16-JUN-1998; 98US-0089406.

XX (NOVA-) NOVA MOLECULAR INC.  
 XX  
 PI Sevigny P, Wiebusch H, Schappert K;  
 XX  
 DR WPI: 2000-126550/11.  
 DR N-PSDB: AA249471.  
 XX  
 PT Prediction of drug efficacy for treating neurological diseases like  
 PT Alzheimer's disease, neurofibromatosis, Huntington's disease -  
 XX  
 PS Disclosure: Fig 4: 37pp: English.  
 XX  
 CC The present sequence is the human polymorphic variant  
 CC of butyrylcholinesterase-K (BChE-K) protein. BChE-K is an allelic variant  
 CC of BChE. Determining BChE allele status (homozygous or heterozygous) of a  
 CC patient helps predicting risk of neurological diseases, efficacy of  
 CC therapy and determining treatment protocol. BChE-K and apoE4 allele  
 CC status also indicate patient's risk for having a neurological disease.  
 CC This method enables treating Alzheimer's disease, Huntington's disease,  
 CC depression, neurofibromatosis, amyotrophic lateral sclerosis, stroke,  
 CC multiple sclerosis, Parkinson's disease, multi-infarct dementia and  
 CC other non-AD neurological diseases.  
 CC  
 SQ Sequence 602 AA;

Query Match 48.4%; Score 44; DB 21; Length 602;  
 Best Local Similarity 46.7%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;  
 QY 3 SRVPTAW--AFHGY 15  
 I::||| | |||  
 DB 454 sklpewmgymhy 468

RESULT 15  
 AA49471  
 ID AA49471 standard; protein; 602 AA.  
 XX  
 AC AA49471;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DE Human wild-type butyryl cholinesterase (BChE).  
 XX  
 KW Organophosphate; detoxification; esterase; acetylcholinesterase; AChE;  
 KW butyrylcholinesterase; BChE; carboxylesterase; CaE; sheep dip; human;  
 KW nerve agent; organophosphorus acid anhydride; OPAA.  
 XX  
 OS Homo sapiens.  
 OS  
 XX US6001625-A.  
 PN  
 XX 14-DEC-1999.  
 PD  
 XX 19-MAY-1995; 95US-0446100.  
 PF  
 XX 19-MAY-1995; 95US-0446100.  
 PR  
 XX (USSA ) US SEC OF ARMY.  
 PA  
 XX Broomfield CA, Lockridge O, Millard CB;  
 PI  
 XX WPI: 2000-096137/08.  
 DR  
 XX  
 PT Enhancing the organophosphate detoxifying capabilities of esterases for  
 PT the treatment of organophosphate poisoning -  
 XX  
 PS Disclosure; Columns 3-4; 64pp: English.  
 XX  
 CC The invention provides a method of enhancing organophosphate detoxifying  
 CC capabilities of esterases (either human acetylcholinesterases (AChE),

CC human butyrylcholinesterases (BChE) and/or carboxylesterases (CaE),  
 CC that comprises substituting a histidine residue for 1 or more amino  
 CC acid(s) within 6 Angstrom of an active site serine. The method may be  
 CC used for enhancing organophosphate detoxifying capabilities of esterases  
 CC (either human AChE, human BChE and/or human CaE). The modified esterases  
 CC may then be used to treat agricultural workers poisoned with  
 CC organophosphates through contact with chemical such as sheep dips. They  
 CC may also be used to treat military personnel contaminated by chemical  
 CC weaponry such as nerve agents. Additionally, the esterases may also be  
 CC used to decontaminate ground and buildings and equipment used to store,  
 CC or contaminated by organophosphates. The method produces esterases with  
 CC improved detoxification properties over naturally occurring  
 CC organophosphorus acid anhydride (OPAA) hydrolyzing enzymes. They are also  
 CC less likely to be inactivated by the OPAA.  
 CC  
 SQ Sequence 602 AA;

Query Match 48.4%; Score 44; DB 21; Length 602;  
 Best Local Similarity 46.7%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;  
 QY 3 SRVPTAW--AFHGY 15  
 I::||| | |||  
 DB 454 sklpewmgymhy 468

Search completed: August 15, 2002, 11:38:24  
 Job time: 265 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:34:34 ; Search time 26.88 Seconds  
(without alignments)  
13.630 Million cell updates/sec

Title: US-09-613-092A-5  
Perfect score: 91  
Sequence: 1 TVSRPWTAWAFHGY 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/CTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	48.4	572	6 5200183-5	Patent No. 5200183
2	44	48.4	573	6 5215909-12	Patent No. 5215909
3	44	48.4	602	3 US-08-446-100-1	Sequence 1, Appl
4	44	48.4	602	3 US-08-446-100-2	Sequence 2, Appl
5	44	48.4	602	3 US-08-446-100-3	Sequence 3, Appl
6	44	48.4	602	3 US-08-446-100-4	Sequence 4, Appl
7	44	48.4	602	3 US-08-446-100-5	Sequence 5, Appl
8	44	48.4	602	3 US-08-446-100-6	Sequence 6, Appl
9	44	48.4	602	3 US-08-446-100-7	Sequence 7, Appl
10	44	48.4	602	3 US-08-446-100-8	Sequence 8, Appl
11	44	48.4	602	3 US-08-446-100-9	Sequence 9, Appl
12	44	48.4	602	3 US-08-446-100-10	Sequence 10, Appl
13	44	48.4	602	3 US-08-446-100-11	Sequence 11, Appl
14	44	48.4	602	3 US-08-446-100-12	Sequence 12, Appl
15	44	48.4	602	3 US-08-446-100-13	Sequence 13, Appl
16	44	48.4	602	3 US-08-446-100-14	Sequence 14, Appl
17	44	48.4	602	3 US-08-446-100-15	Sequence 15, Appl
18	44	48.4	602	3 US-08-446-100-16	Sequence 16, Appl
19	44	48.4	602	3 US-08-446-100-17	Sequence 17, Appl
20	44	48.4	602	3 US-08-446-100-18	Sequence 18, Appl
21	44	48.4	602	3 US-08-446-100-19	Sequence 19, Appl
22	44	48.4	602	4 US-09-334-489-3	Sequence 24, Appl
23	44	48.4	602	4 US-09-334-489-4	Sequence 3, Appl
24	44	48.4	602	4 US-09-334-489-5	Sequence 4, Appl
25	44	48.4	635	6 5215909-10	Patent No. 5215909
26	42	46.2	171	4 US-08-611-587-5	Sequence 5, Appl
27	42	46.2	761	1 US-07-906-395-2	Sequence 2, Appl

28	42	46.2	761	1 US-08-192-632-2	Sequence 2, Appl
29	42	46.2	761	1 US-08-710-676-2	Sequence 2, Appl
30	42	46.2	761	3 US-09-099-902B-2	Sequence 2, Appl
31	42	46.2	761	5 PCT-US93-06080-2	Sequence 2, Appl
32	40	44.0	150	4 US-08-334-179A-10	Sequence 10, Appl
33	40	44.0	1038	4 US-08-334-179A-8	Sequence 8, Appl
34	39	42.9	53	2 US-08-799-173A-13	Sequence 13, Appl
35	39	42.9	55	1 US-07-862-021B-20	Sequence 20, Appl
36	39	42.9	55	5 PCT-US93-03164-20	Sequence 20, Appl
37	39	42.9	807	1 US-07-862-021B-10	Sequence 10, Appl
38	39	42.9	807	1 US-08-313-288B-10	Sequence 10, Appl
39	39	42.9	807	5 PCT-US93-03164-10	Sequence 10, Appl
40	39	42.9	984	4 US-09-287-354-2	Sequence 2, Appl
41	39	42.9	1189	4 US-09-287-354-3	Sequence 3, Appl
42	39	42.9	1189	4 US-09-287-354-4	Sequence 4, Appl
43	38	41.8	22	1 US-08-178-708-2	Sequence 2, Appl
44	38	41.8	22	1 US-08-457-552-2	Sequence 2, Appl
45	38	41.8	22	1 US-08-456-430-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
5200183-5  
; Patent No. 5200183  
; APPLICANT: TANG, JORDAN J.N.; RANG, CHI-SUN  
; TITLE OF INVENTION: RECOMBINANT BLUE SALT ACTIVATED LIPASES  
; NUMBER OF SEQUENCES: 22  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,426  
; FILING DATE: 12-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 504,635  
; FILING DATE: 04-APR-1990  
; APPLICATION NUMBER: 122,410  
; FILING DATE: 19-NOV-1987  
; SEQ ID NO:5  
; LENGTH: 572  
5200183-5

Query Match 48.4%; Score 44; DB 6; Length 572;  
Best Local Similarity 46.7%; Pred. No. 53;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPWTAW-AFHGY 15  
Db 424 SKLPWPMGVMHGY 438

RESULT 2  
5215909-12  
; Patent No. 5215909  
; APPLICANT: SOREO, HERMONA  
; TITLE OF INVENTION: HUMAN CHOLINESTERASE GENES  
; NUMBER OF SEQUENCES: 13  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/572,911  
; FILING DATE: 15-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 87,724  
; FILING DATE: 21-AUG-1987  
; APPLICATION NUMBER: 875,737  
; FILING DATE: 18-JUN-1986  
; SEQ ID NO:12  
; LENGTH: 573  
5215909-12

Query Match 48.4%; Score 44; DB 6; Length 573;  
Best Local Similarity 46.7%; Pred. No. 53;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;



```

; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: broomfield
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 602 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: YES
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: human esterases
; US-08-446-100-3

Query Match 48.4%; Score 44; DB 3; Length 602;
Best Local Similarity 46.7%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPTAW--AFHGY 15
Db 454 SKLPPEMGMVHGY 468

RESULT 6
US-08-446-100-4
; Sequence 4, Application US/08446100
; Patent No. 6001625
; GENERAL INFORMATION:
; APPLICANT: Broomfield, Clarence A
; APPLICANT: Millard, Charles B
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Site-directed Mutagenesis of Esterases
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Assoc.
; STREET: 9669 A Main Street, P.O. Box 2509
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,100
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: broomfield
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 602 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: YES
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: human esterases
; US-08-446-100-5
```

```

; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: human esterases
; US-08-446-100-4

Query Match 48.4%; Score 44; DB 3; Length 602;
Best Local Similarity 46.7%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPTAW--AFHGY 15
Db 454 SKLPPEMGMVHGY 468

RESULT 7
US-08-446-100-5
; Sequence 5, Application US/08446100
; Patent No. 6001625
; GENERAL INFORMATION:
; APPLICANT: Broomfield, Clarence A
; APPLICANT: Millard, Charles B
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Site-directed Mutagenesis of Esterases
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Assoc.
; STREET: 9669 A Main Street, P.O. Box 2509
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,100
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: broomfield
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 602 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: YES
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: human esterases
; US-08-446-100-5

Query Match 48.4%; Score 44; DB 3; Length 602;
Best Local Similarity 46.7%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPTAW--AFHGY 15
Db 454 SKLPPEMGMVHGY 468

RESULT 8
```

```
US-08-446-100-6
; Sequence 6, Application US/08446100
; Patent No. 6001625
; GENERAL INFORMATION:
; APPLICANT: Broomfield, Clarence A
; APPLICANT: Millard, Charles B
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Assoc.
; STREET: 9669 A Main Street, P.O. Box 2509
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,100
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: Broomfield
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 602 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: YES
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: human esterases
;
US-08-446-100-6

Query Match      48.4%; Score 44; DB 3; Length 602;
Best Local Similarity 46.7%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY      3 SRVPTAW-AFHGY 15
      |::||| | |||
Db      454 SKLPPEWGMGYHG 468

RESULT 9
US-08-446-100-7
; Sequence 7, Application US/08446100
; Patent No. 6001625
; GENERAL INFORMATION:
; APPLICANT: Broomfield, Clarence A
; APPLICANT: Millard, Charles B
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Assoc.
; STREET: 9669 A Main Street, P.O. Box 2509
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
```

```
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,100
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: Broomfield
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 602 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: YES
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: human esterases
;
US-08-446-100-7

Query Match      48.4%; Score 44; DB 3; Length 602;
Best Local Similarity 46.7%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY      3 SRVPTAW-AFHGY 15
      |::||| | |||
Db      454 SKLPPEWGMGYHG 468

RESULT 10
US-08-446-100-8
; Sequence 8, Application US/08446100
; Patent No. 6001625
; GENERAL INFORMATION:
; APPLICANT: Broomfield, Clarence A
; APPLICANT: Millard, Charles B
; APPLICANT: Lockridge, Oksana
; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Assoc.
; STREET: 9669 A Main Street, P.O. Box 2509
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,100
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: Broomfield
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
```

INFORMATION FOR SEQ ID NO: 8;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 602 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: YES  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
US-08-446-100-8  
ORGANISM: human esterases

Query Match 48.4%; Score 44; DB 3; Length 602;  
Best Local Similarity 46.7%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 3 SRVPTAW--AFHGY 15  
|:|:| | |  
Db 454 SKLPWEMMGVNHGY 468

RESULT 11  
US-08-446-100-9  
; Sequence 9, Application US/08446100  
; Patent No. 6001625  
; GENERAL INFORMATION:  
; APPLICANT: Broomfield, Clarence A  
; APPLICANT: Millard, Charles B  
; APPLICANT: Lockridge, Oksana  
; TITLE OF INVENTION: Site-directed mutagenesis of Esterases  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hendricks and Assoc.  
; STREET: 9669 A Main Street, P.O. Box 2509  
; CITY: Fairfax  
; STATE: VA  
; COUNTRY: US  
; ZIP: 22031  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,100  
; FILING DATE: 19-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hendricks, Glenna  
; REGISTRATION NUMBER: 32,535  
; REFERENCE/DOCKET NUMBER: broomfield  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 425-4250  
; TELEFAX: (703) 425-2767  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 602 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: YES  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: human esterases  
US-08-446-100-9

Query Match 48.4%; Score 44; DB 3; Length 602;

Best Local Similarity 46.7%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 3 SRVPTAW--AFHGY 15  
|:|:| | |  
Db 454 SKLPWEMMGVNHGY 468

RESULT 12  
US-08-446-100-10  
; Sequence 10, Application US/08446100  
; Patent No. 6001625  
; GENERAL INFORMATION:  
; APPLICANT: Broomfield, Clarence A  
; APPLICANT: Millard, Charles B  
; APPLICANT: Lockridge, Oksana  
; TITLE OF INVENTION: Site-directed mutagenesis of Esterases  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hendricks and Assoc.  
; STREET: 9669 A Main Street, P.O. Box 2509  
; CITY: Fairfax  
; STATE: VA  
; COUNTRY: US  
; ZIP: 22031  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,100  
; FILING DATE: 19-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hendricks, Glenna  
; REGISTRATION NUMBER: 32,535  
; REFERENCE/DOCKET NUMBER: broomfield  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 425-4250  
; TELEFAX: (703) 425-2767  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 602 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: YES  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: human esterases  
US-08-446-100-10

Query Match 48.4%; Score 44; DB 3; Length 602;  
Best Local Similarity 46.7%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 3 SRVPTAW--AFHGY 15  
|:~|:~| | |  
Db 454 SKLPWEMMGVNHGY 468

RESULT 13  
US-08-446-100-11  
; Sequence 11, Application US/08446100  
; Patent No. 6001625  
; GENERAL INFORMATION:  
; APPLICANT: Broomfield, Clarence A  
; APPLICANT: Millard, Charles B  
; APPLICANT: Lockridge, Oksana

;; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases  
;; NUMBER OF SEQUENCES: 31  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Hendricks and Assoc.  
;; STREET: 9669 A Main Street, P.O. Box 2509  
;; CITY: Fairfax  
;; STATE: VA  
;; COUNTRY: US  
;; ZIP: 22031  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/446,100  
;; FILING DATE: 19-MAY-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hendricks, Glenna  
;; REGISTRATION NUMBER: 32,535  
;; REFERENCE/DOCKET NUMBER: broomfield  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 425-4250  
;; TELEFAX: (703) 425-2767  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 602 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: YES  
;; ANTI-SENSE: YES  
;; FRAGMENT TYPE: N-terminal  
;; ORIGINAL SOURCE:  
;; ORGANISM: human esterases  
;; US-08-446-100-11

Query Match 48.4%; Score 44; DB 3; Length 602;  
Best Local Similarity 46.7%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPTAW-AFHGY 15  
|::||| |  
Db 454 SKLPWMGYMHGY 468

RESULT 14  
US-08-446-100-12  
;; Sequence 12, Application US/08446100  
;; Patent No. 6001625  
;; GENERAL INFORMATION:  
;; APPLICANT: Broomfield, Clarence A  
;; APPLICANT: Millard, Charles B  
;; APPLICANT: Lockridge, Oksana  
;; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases  
;; NUMBER OF SEQUENCES: 31  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Hendricks and Assoc.  
;; STREET: 9669 A Main Street, P.O. Box 2509  
;; CITY: Fairfax  
;; STATE: VA  
;; COUNTRY: US  
;; ZIP: 22031  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/446,100

;; FILING DATE: 19-MAY-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hendricks, Glenna  
;; REGISTRATION NUMBER: 32,535  
;; REFERENCE/DOCKET NUMBER: broomfield  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 425-4250  
;; TELEFAX: (703) 425-2767  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 602 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: YES  
;; ANTI-SENSE: YES  
;; FRAGMENT TYPE: N-terminal  
;; ORIGINAL SOURCE:  
;; ORGANISM: human esterases  
;; US-08-446-100-12

Query Match 48.4%; Score 44; DB 3; Length 602;  
Best Local Similarity 46.7%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPTAW-AFHGY 15  
|::||| |  
Db 454 SKLPWMGYMHGY 468

RESULT 15  
US-08-446-100-13  
;; Sequence 13, Application US/08446100  
;; Patent No. 6001625  
;; GENERAL INFORMATION:  
;; APPLICANT: Broomfield, Clarence A  
;; APPLICANT: Millard, Charles B  
;; APPLICANT: Lockridge, Oksana  
;; TITLE OF INVENTION: Site-Directed Mutagenesis of Esterases  
;; NUMBER OF SEQUENCES: 31  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Hendricks and Assoc.  
;; STREET: 9669 A Main Street, P.O. Box 2509  
;; CITY: Fairfax  
;; STATE: VA  
;; COUNTRY: US  
;; ZIP: 22031  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/446,100  
;; FILING DATE: 19-MAY-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hendricks, Glenna  
;; REGISTRATION NUMBER: 32,535  
;; REFERENCE/DOCKET NUMBER: broomfield  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 425-4250  
;; TELEFAX: (703) 425-2767  
;; INFORMATION FOR SEQ ID NO: 13:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 602 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein

; HYPOTHETICAL: YES  
 ; ANTI-SENSE: YES  
 ; FRAGMENT TYPE: N-terminal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: human esterases  
 US-08-446-100-13

Query Match 48.4%; Score 44; DB 3; Length 602;  
 Best Local Similarity 46.7%; Pred. No. 56;  
 Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

Qy 3 SRVPTAW--AFHG 15  
 I::I I I I I  
 Db 454 SKLPPEWNGVMHGY 468

Search completed: August 15, 2002, 11:38:58  
 Job time: 264 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:36:19 ; Search time 35.58 Seconds  
(without alignments)  
40.510 Million cell updates/sec

Title: US-09-613-092a-5  
Perfect score: 91  
Sequence: 1 TVSRVPMWTAMAFHG 15

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR-71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	49.5	278	2	hypotheetical prote
2	45	49.5	454	2	F9L1.4 protein - A
3	45	49.5	537	2	hypotheetical prote
4	44	48.4	350	2	probable membrane
5	44	48.4	581	2	cholinesterase (EC
6	44	48.4	602	1	cholinesterase (EC
7	44	48.4	603	2	cholinesterase (EC
8	43.5	47.8	310	2	hypotheetical prote
9	43.5	47.8	457	2	probable NAM-1-like
10	43	47.3	302	2	hypotheetical prote
11	43	47.3	351	2	hypotheetical prote
12	43	47.3	441	2	hypotheetical prote
13	43	47.3	505	2	probable two compo
14	43	47.3	505	2	sensor histidine k
15	43	47.3	658	2	endoglucanase 3 (E
16	43	47.3	722	2	hypotheetical prote
17	43	47.3	1120	2	hypotheetical prote
18	42	46.2	142	2	DNM-directed RNA p
19	42	46.2	247	2	tapeum-specific p
20	42	46.2	297	2	conserved hypothe
21	42	46.2	299	2	hypotheetical prote
22	42	46.2	317	2	hypotheetical prote
23	42	46.2	650	2	hypotheetical prote
24	42	46.2	651	2	conserved amine oxi
25	42	46.2	656	2	hypotheetical prote
26	42	46.2	656	2	hypotheetical prote
27	42	46.2	666	1	MG032 homolog B01_
28	42	46.2	750	2	cation-transportin
29	42	46.2	750	2	cation-transportin

30	42	46.2	752	2	probable ctbp prot
31	42	46.2	761	2	probable ctbp prot
32	42	46.2	766	2	probable cation-tr
33	42	46.2	780	2	cation-transportin
34	41	45.1	232	2	trbp protein - Ent
35	41	45.1	461	2	hypotheetical prote
36	41	45.1	543	2	hypotheetical prote
37	41	45.1	602	2	acetylcholinestera
38	41	45.1	1263	2	heterocyst glycoli
39	40	44.0	155	2	hypotheetical prote
40	40	44.0	155	2	hypotheetical prote
41	40	44.0	200	2	hypotheetical prote
42	40	44.0	242	2	homeotic protein G
43	40	44.0	277	2	protein F41H0.5 [
44	40	44.0	353	2	hypotheetical prote
45	40	44.0	379	2	hypotheetical prote

## ALIGNMENTS

RESULT 1  
S46681  
hypotheetical protein YHR192w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypotheetical protein H9998.8  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 23-Mar-2001  
C:Accession: S46681  
R:Mac1, C.  
submitted to the EMBL Data Library, February 1994  
A:Description: The sequence of S. cerevisiae cosmid 9998.  
A:Reference number: S46674  
A:Accession: S46681  
A:Molecule type: DNA  
A:Residues: 1-278 <MAC>  
A:Cross-references: EMBL:U00030; NID:g458927; PIDN:AAB68359.1; PID:g458933; MIPS:YHR1  
C:Genetics:  
A:Map position: 8R  
C:Keywords: transmembrane protein

Query Match 49.5%; Score 45; DB 2; Length 278;  
Best Local Similarity 46.7%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 TVSRVPMWTAMAFHG 15  
|| : : | : |||  
DB 62 TVLALSYTYMEYHG 76

RESULT 2  
G86284  
F9L1.4 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
C:Accession: G86284  
R:Thologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: G86284  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <STO>  
A:Cross-references: GB:AE005172; NID:g5103809; PIDN:AAB39639.1; GSPDB:GN00141  
C:Genetics:

A;Map position: 1

Query Match 49.5%; Score 45; DB 2; Length 454;  
 Best Local Similarity 85.7%; Pred. No. 32;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PWTAMAF 12  
 |||||  
 Db 32 PWTAMAY 38

RESULT 3

hypothetical protein F13112.30 - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C;Accession: T45634  
 R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; Sa  
 submitted to the Protein Sequence Database, November 1999  
 A;Reference number: Z23010  
 A;Accession: T45634  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1537 <CHO>  
 A;Cross-references: EMBL:AL133292  
 A;Experimental source: cultivar Columbia; BAC clone F13112  
 C;Genetics:  
 A;Map position: 3  
 A;Introns: 35/2; 153/3; 229/2; 350/3; 399/3; 445/3; 457/1  
 A;Note: F13112.30

Query Match 49.5%; Score 45; DB 2; Length 537;  
 Best Local Similarity 61.5%; Pred. No. 37;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 SRPWTAMAFHG 15  
 |||||  
 Db 363 SAVPMSMAFRTG 375

RESULT 4

probable membrane protein YNL264c - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: hypothetical protein N0815  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 20-Jun-2000  
 C;Accession: S60917; S63237; S65119  
 R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.M.  
 submitted to the EMBL Data Library, October 1995  
 A;Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV fr  
 A;Reference number: S60909  
 A;Accession: S60917  
 A;Molecule type: DNA  
 A;Residues: 1-350 <SPN>  
 A;Cross-references: EMBL:X02494; NID:g1045236; PIDN:CAA63233.1; PID:g1045245  
 R;Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.  
 submitted to the Protein Sequence Database, April 1996  
 A;Reference number: S63235  
 A;Accession: S63237  
 A;Molecule type: DNA  
 A;Residues: 1-350 <SEM>  
 A;Cross-references: EMBL:Z71540; NID:g1302320; PIDN:CAA96171.1; PID:g1302321; MIPS:YNL26  
 A;Experimental source: strain S288C  
 R;Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.  
 Yeast 12, 505-514, 1996  
 A;Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from Sa  
 A;Reference number: S65119; MUID:96310631  
 A;Accession: S65119  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-350 <SEF>

A;Cross-references: EMBL:X02494; NID:g1045236; PIDN:CAA63233.1; PID:g1045245  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
 C;Genetics:  
 A;Map position: 14L  
 A;Note: YNL264  
 C;Superfamily: probable membrane protein YNL264c; cellular retinaldehyde-binding prot  
 C;Keywords: transmembrane protein  
 F;96-291/Domain: cellular retinaldehyde-binding protein homology <CRB>  
 F;239-255/Domain: transmembrane #status predicted <TM>

Query Match 48.4%; Score 44; DB 2; Length 350;  
 Best Local Similarity 75.0%; Pred. No. 35;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 VPWTAMAF 12  
 |||||  
 Db 247 IPWTAMAF 254

RESULT 5

cholinesterase (EC 3.1.1.8) - rabbit  
 N;Alternate names: butyrylcholinesterase  
 C;Species: Oryctolagus cuniculus (domestic rabbit)  
 C;Date: 14-Feb-1992 #sequence\_revision 01-Mar-1996 #text\_change 20-Jun-2000  
 C;Accession: S10255; C39768  
 R;Jbilo, O.; Chatonnet, A.  
 Nucleic Acids Res. 18, 3990, 1990  
 A;Title: Complete sequence of rabbit butyrylcholinesterase.  
 A;Reference number: S10255; MUID:90326526  
 A;Accession: S10255  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-581 <JBI>  
 A;Cross-references: EMBL:X52090; NID:g1476; PIDN:CAA36308.1; PID:g1370277  
 R;Appagaus, M.; Chatonnet, A.; Masson, P.; Newton, M.; Vaughan, T.A.; Bartels, C.F.;  
 J. Biol. Chem. 266, 6966-6974, 1991  
 A;Title: Use of the polymerase chain reaction for homology probing of butyrylcholines  
 A;Reference number: A39768; MUID:91201348  
 A;Accession: C39768  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 75-215 <ARP>  
 A;Cross-references: GB:M62779; NID:g164788; PIDN:AAA31169.1; PID:g164789  
 C;Genetics:  
 A;Introns: 485/2; 541/1  
 C;Superfamily: cholinesterase; cholinesterase homology  
 C;Keywords: carboxylic ester hydrolase; glycoprotein  
 F;35-535/Domain: cholinesterase homology <CHE>

Query Match 48.4%; Score 44; DB 2; Length 581;  
 Best Local Similarity 46.7%; Pred. No. 56;  
 Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRPWTAM--AFHG 15  
 |||||  
 Db 433 SKLPPEWGMVHG 447

RESULT 6

ACHU  
 cholinesterase (EC 3.1.1.8) precursor [validated] - human  
 N;Alternate names: acylcholine acylhydrolase; butyrylcholinesterase; choline esterase  
 C;Species: Homo sapiens (man)  
 C;Date: 30-Jun-1987 #sequence\_revision 23-Feb-1996 #text\_change 08-Dec-2000  
 C;Accession: A33769; A26613; A33887; A34668; A00772  
 R;Appagaus, M.; Kott, M.; Vatsis, K.P.; Bartels, C.F.; La Du, B.N.; Lockridge, O.  
 Biochemistry 29, 124-131, 1990  
 A;Title: Structure of the gene for human butyrylcholinesterase. Evidence for a single  
 A;Reference number: A33769; MUID:90212557  
 A;Accession: A33769

A: Molecule type: DNA  
 A: Residues: 'MSVQSNLQAGAAASCISPKYYMIFTPCKLCHLCCRESEIN', 1-602 <ARP>  
 A: Cross-references: GB:M32391; GB:J02879  
 A: Note: two ATG codons found upstream of Met-1 do not lie in a favorable context for translation.  
 R: Prody, C.A.; Zevin-Sonkin, D.; Gnatli, A.; Goldberg, O.; Soreq, H.  
 C: Species: Mus musculus (house mouse)  
 C: Date: 28-Oct-1996 #sequence\_revision 08-Nov-1996 #text\_change 18-Jun-1999  
 C: Accession: S70849; S15680; A39768  
 R: Taylor, P.  
 Submitted to the EMBL Data Library, August 1992  
 A: Reference number: S70849  
 A: Accession: S70849

A: Molecule type: DNA  
 A: Residues: 1-133, 'D', 135-602 <PRO>  
 R: McLeenan, C.; Adkins, S.; Chalmers, A.; Vaughan, T.A.; Bartels, C.F.; Kott, M.; Rose  
 Proc. Natl. Acad. Sci. U.S.A. 84, 3555-3559, 1987  
 A: Title: Brain cDNA clone for human cholinesterase.  
 A: Reference number: A33887; MUID:88016155  
 A: Accession: A33887

A: Molecule type: mRNA  
 A: Residues: 'MSVQSNLQAGAAASCISPKYYMIFTPCKLCHLCCRESEIN', 1-602 <MCT>  
 A: Note: two ATG codons found upstream of Met-1 do not lie in a favorable context for translation.  
 R: Nozue, C.P.; McGuire, M.C.; Graesser, C.; Bartels, C.F.; Arpagaus, M.; Van der Spek, A.M. J. Hum. Genet. 46, 934-942, 1990  
 A: Title: Identification of a frameshift mutation responsible for the silent phenotype of A: Reference number: A34668; MUID:90252779  
 A: Accession: A34668

A: Molecule type: DNA  
 A: Residues: 143-145, 'VSNMNIIFTC', <NOG>  
 A: Note: frameshift mutant in codon for residue 145 (Gly)  
 R: Lockridge, O.; Bartels, C.F.; Vaughan, T.A.; Wong, C.K.; Norton, S.E.; Johnson, L.L.  
 J. Biol. Chem. 265, 549-557, 1987  
 A: Title: Complete amino acid sequence of human serum cholinesterase.  
 A: Reference number: A00772; MUID:87109144  
 A: Accession: A00772

A: Molecule type: protein  
 A: Residues: 29-602 <LOC>  
 A: Experimental source: plasma  
 C: Comment: Cholinesterase is present in most cells (except erythrocytes).  
 C: Genetics:  
 A: Gene: GDB:BCH; CHE1  
 A: Cross-references: GDB:120558; OMIM:177400  
 A: Map position: 3q26.1-3q26.2  
 A: Introns: 506/2; 562/1  
 C: Function:  
 A: Description: hydrolyzes acylcholines to choline and a carboxylic acid  
 A: Note: this cholinesterase is highly reactive with organophosphate esters  
 C: Superfamily: cholinesterase; cholinesterase homology  
 C: Keywords: carboxylic ester hydrolase; glycoprotein; homotetramer  
 F: 1-28/Domain: signal sequence #status predicted <SIG>  
 F: 29-602/Product: cholinesterase #status experimental <MAT>  
 F: 56-556/Domain: cholinesterase homology <CHE>  
 F: 45,85,134,269,284,359,483,509,514/Binding site: carbohydrate (asn) (covalent) #status F:226/Active site: Ser #status experimental

Query Match 48.4%; Score 44; DB 1; Length 602;  
 Best Local Similarity 46.7%; Pred. No. 58;  
 Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPTAW--AFHGY 15  
 |::|||  
 Db 454 SKLPWPMGVWVHGY 468

RESULT 7  
 S70849  
 N: Alternate names: butyrylcholine esterase  
 C: Species: Mus musculus (house mouse)  
 C: Date: 28-Oct-1996 #sequence\_revision 08-Nov-1996 #text\_change 18-Jun-1999  
 C: Accession: S70849; S15680; A39768  
 R: Taylor, P.  
 Submitted to the EMBL Data Library, August 1992  
 A: Reference number: S70849  
 A: Accession: S70849

A: Molecule type: nucleic acid  
 A: Residues: 1-603 <TAY>  
 A: Cross-references: EMBL:M99492; NID:G191579; PIDN:AAA7328.1; PID:G191580  
 R: Rachalsky, T.L.; Camp, S.; Li, Y.; Ekstrom, T.J.; Newton, M.; Taylor, P.  
 Neuron 5, 317-327, 1990  
 A: Title: Molecular cloning of mouse acetylcholinesterase: tissue distribution of alternative reference number: JH0314; MUID:90380429  
 A: Accession: S15680

A: Status: nucleic acid sequence not shown  
 A: Molecule type: nucleic acid  
 A: Residues: 30-128, 'P', 130-603 <RAC>  
 A: Cross-references: EMBL:M99492  
 R: Arpagaus, M.; Chalmers, A.; Masson, P.; Newton, M.; Vaughan, T.A.; Bartels, C.F.;  
 J. Biol. Chem. 265, 6966-6974, 1991  
 A: Title: Use of the polymerase chain reaction for homology probing of butyrylcholines  
 A: Reference number: A39768; MUID:91201348  
 A: Accession: A39768

A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 97-128, 'P', 130-237 <ARP>  
 C: Superfamily: cholinesterase; cholinesterase homology  
 C: Keywords: carboxylic ester hydrolase; glycoprotein  
 F: 57-557/Domain: cholinesterase homology <CHE>

Query Match 48.4%; Score 44; DB 2; Length 603;  
 Best Local Similarity 46.7%; Pred. No. 58;  
 Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 3 SRVPTAW--AFHGY 15  
 |::|||  
 Db 455 SKLPWPMGVWVHGY 469

RESULT 8  
 T01939  
 hypothetical protein F1104.5 - Arabidopsis thaliana  
 C: Species: Arabidopsis thaliana (mouse-ear cress)  
 C: Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999  
 C: Accession: T01939  
 R: Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevasakis, E.  
 Submitted to the EMBL Data Library, October 1998  
 A: Description: The sequence of A. thaliana F1104.  
 A: Reference number: Z14466  
 A: Accession: T01939

A: Status: translated from GR/EMBL/DBJ  
 A: Molecule type: DNA  
 A: Residues: 1-310 <ABU>  
 A: Cross-references: EMBL:AF096370; NID:G3695372; PID:G3695378  
 A: Experimental source: cultivar Columbia  
 C: Genetics:  
 A: Map position: 4  
 A: Introns: 28/1; 118/3; 165/3  
 A: Note: F1104.5

Query Match 47.8%; Score 43.5; DB 2; Length 310;  
 Best Local Similarity 53.3%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 4 RVP--WTAWAFHGY 15  
 ||| |||  
 Db 95 RVPKGVWTPWVWVHGY 109

RESULT 9  
 B85020  
 probable NAM-like protein [imported] - Arabidopsis thaliana  
 C: Species: Arabidopsis thaliana (mouse-ear cress)  
 C: Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C: Accession: B85020  
 R: Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp  
 Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: AB5001; MUID:20083488  
A:Accession: B85020  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-457 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7268198; PIDN:CAB77725.1; GSPDB:GNC0140  
C:Genetics:  
A:Gene: AT4G01550  
A:Map position: 4

Query Match 47.8%; Score 43.5; DB 2; Length 457;  
Best Local Similarity 53.3%; Pred. No. 53;  
Matches 8; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 4 RVP--WTAMAFHG 15  
||| ||| |||  
Db 119 RVRGVTWPMVMEY 133

RESULT 10  
S50579  
hypothetical protein YER076c - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 23-Mar-2001  
C:Accession: S50579  
R:Dietrich, F.S.  
Submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmids 9747, 8198, 9781, and lambda clones  
A:Reference number: S50436  
A:Accession: S50579  
A:Molecule type: DNA  
A:Residues: 1-302 <DIE>  
A:Cross-references: EMBL:U18839; NID:g603313; PID:g603314; GSPDB:GNC00005; MIPS:YER076c  
C:Genetics:  
A:Gene: MIPS:YER076c  
A:Map position: 5R

Query Match 47.3%; Score 43; DB 2; Length 302;  
Best Local Similarity 75.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 WTAMAFHG 14  
||| ||| ||  
Db 180 WVAWAVHG 187

RESULT 11  
S63197  
hypothetical protein YNL231c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein N1158  
C:Species: Saccharomyces cerevisiae  
C:Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 20-Jun-2000  
C:Accession: S63197; S67367; S72085  
R:Pandolfo, D.; De Antoni, A.; Lanfranchi, G.; Valle, G.  
Submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63188  
A:Accession: S63197  
A:Molecule type: DNA  
A:Residues: 1-351 <PAN>  
A:Cross-references: EMBL:Z71507; NID:g1302256; PIDN:CAA96136.1; PID:g1302257; MIPS:YNL231  
A:Experimental source: strain S288C  
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.  
Submitted to the EMBL Data Library, February 1996  
A:Description: DNA sequence of cosmid 14-5 from chromosome XIV.  
A:Reference number: S67355  
A:Accession: S67367  
A:Molecule type: DNA  
A:Residues: 1-351 <PAW>  
A:Cross-references: EMBL:Z69381; NID:g1183970; PIDN:CAA93367.1; PID:g1183983  
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.

Yeast 12, 1071-1076, 1996  
A:Title: The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open reading  
A:Reference number: S72073; MUID:97051596  
A:Accession: S72085  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-351 <PAF>  
A:Cross-references: EMBL:Z69381; NID:g1183970; PIDN:CAA93367.1; PID:g1183983  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996  
C:Genetics:  
A:Map position: 14L  
A:Note: YNL231c  
C:Superfamily: probable membrane protein YNL264c; cellular retinaldehyde-binding prot  
F:g1-289/Domain: cellular retinaldehyde-binding protein homology <CRB>

Query Match 47.3%; Score 43; DB 2; Length 351;  
Best Local Similarity 45.5%; Pred. No. 49;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSRVPTWMAF 12  
:: ||| |||  
Db 242 LTNPMLATWF 252

RESULT 12  
G83212  
hypothetical protein PA3464 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83212  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: AB2950; MUID:20437337  
A:Accession: G83212  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-441 <STO>  
A:Cross-references: GB:AE004767; GB:AE004091; NID:g9949604; PIDN:AA606852.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3464

Query Match 47.3%; Score 43; DB 2; Length 441;  
Best Local Similarity 85.7%; Pred. No. 60;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 PWTAMAF 12  
||||| |  
Db 144 PWTAMRF 150

RESULT 13  
AB1806  
probable two component sensor kinase NMA1803 [imported] - Neisseria meningitidis (str  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: AB1806  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
.; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: AB1775; MUID:20222556  
A:Accession: AB1806  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-505 <PAR>  
A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85029.1; PID:g738  
A:Experimental source: serogroup A, strain Z2491

C:Genetics:  
A:Gene: NMA1803

Query Match 47.3%; Score 43; DB 2; Length 505;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 7 WTMAAFH 13  
|||||:  
DB 51 WTMAAFY 57

RESULT 14  
H81064

sensor histidine kinase NMB1606 [imported] - Neisseria meningitidis (strain MC58 serogroup C)  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: H81064  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; et al., H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; et al.  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: H81064  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-505 <TEXT>  
A:Cross-references: GB:AE002511; GB:AE002098; NID:g7226857; PIDN:AAF41958.1; PID:g722685.  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1606

Query Match 47.3%; Score 43; DB 2; Length 505;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 7 WTMAAFH 13  
|||||:  
DB 51 WTMAAFY 57

RESULT 15  
A33598

endoglucanase 3 (EC 3.2.1.-) - Fibrobacter succinogenes (strain S85)  
C:Species: Fibrobacter succinogenes  
C:Date: 27-Feb-1990 #sequence\_revision 27-Feb-1990 #text\_change 08-Oct-1999  
C:Accession: A33598  
R:McGavin, M.J.; Forsberg, C.W.; Crosby, B.; Bell, A.W.; Dignard, D.; Thomas, D.Y. J. Bacteriol. 171, 5587-5595, 1989  
A:Title: Structure of the cel-3 gene from Fibrobacter succinogenes S85 and characterization of the gene product.  
A:Reference number: A33598; MUID:9008798  
A:Accession: A33598  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-658 <MCG>  
A:Cross-references: GB:M29047; NID:g148570; PIDN:AA24893.1; PID:g148571; GB:M29681  
C:Keywords: glycosidase; hydrolase

Query Match 47.3%; Score 43; DB 2; Length 658;  
Best Local Similarity 45.5%; Pred. No. 88;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 RVPWTAMAFHG 14  
::||| | : |  
DB 625 QIPWTHMGYTG 635

Search completed: August 15, 2002, 11:39:40  
Job time: 201 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:39:04 ; Search time 18.67 seconds

(without alignments)  
31.108 Million cell updates/sec

Title: US-09-613-092a-5  
Perfect score: 91  
Sequence: 1 TVSRVPTWTAAPHGY 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Swissprot\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	45	49.5	278 1	YH22_YEAST
2	44	48.4	350 1	YN04_YEAST
3	44	48.4	581 1	CHLE_YABIT
4	44	48.4	602 1	CHLE_HUMAN
5	44	48.4	603 1	CHLE_MOUSE
6	43	47.3	176 1	VPI_BPP2
7	43	47.3	302 1	YEO6_YEAST
8	43	47.3	351 1	YXN1_YEAST
9	43	47.3	658 1	GUN3_FIBSU
10	43	47.3	1120 1	RPOK_SCHPO
11	42	46.2	299 1	YOG8_YEAST
12	42	46.2	666 1	YAS5_MYCPN
13	42	46.2	750 1	CTPB_MYCLE
14	42	46.2	752 1	CTPB_MYCTU
15	42	46.2	761 1	CTPA_MYCTU
16	42	46.2	780 1	CTPA_MYCLE
17	40	44.0	200 1	YEO1_MYCTU
18	40	44.0	353 1	YKPI_YEAST
19	40	44.0	379 1	YISP_BACCS
20	40	44.0	541 1	BBR1_SCHCO
21	40	44.0	616 1	REF5_HUMAN
22	40	44.0	660 1	CCMF_PSEFL
23	40	44.0	1038 1	BMR2_MOUSE
24	40	44.0	3386 1	POLG_DEN2P
25	40	44.0	3388 1	POLG_DEN2P
26	40	44.0	3390 1	POLG_DEN3
27	40	44.0	3391 1	POLG_DEN26
28	40	44.0	3391 1	POLG_DEN27
29	40	44.0	3391 1	POLG_DEN28
30	40	44.0	3391 1	POLG_DEN29
31	40	44.0	3396 1	POLG_DEN1S
32	39	42.9	395 1	SYT8_MOUSE
33	39	42.9	402 1	ROO_DESGI

34	39	42.9	419 1	MANA_PSEFL	P49424 pseudomonas
35	39	42.9	426 1	PYRC_THEAO	P96081 thermus aqu
36	39	42.9	497 1	BCHY_RHOCA	P26178 rhodobacter
37	39	42.9	807 1	FSPQ_RAT	P35446 rattus norv
38	39	42.9	1180 1	TYR2_MOUSE	O9f117 mus musculu
39	39	42.9	1189 1	HAIR_HUMAN	O43593 homo sapien
40	39	42.3	775 1	MGD1_MOUSE	O9qyh6 mus musculu
41	38.5	42.3	775 1	MGD1_MOUSE	O9qyh6 mus musculu
42	38.5	42.3	778 1	MGD1_HUMAN	O9qyh6 mus musculu
43	38.5	42.3	1327 1	YK06_MYCTU	O10850 mycobacteri
44	38	41.8	207 1	CAT_CAMCO	P22782 campylobact
45	38	41.8	257 1	FOL1_HUMAN	P15328 homo sapien

## ALIGNMENTS

RESULT 1	YH22_YEAST	STANDARD:	PRT:	278 AA.
ID	YH22_YEAST			
AC	P38878:			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DE	01-OCT-1996 (Rel. 34, Last annotation update)			
GN	Hypothetical 32.1 kDa protein in CTF8-EGD2 intergenic region.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=5288C / AB972;			
RX	MEDLINE=94378003; PubMed=8091229;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,			
RA	Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,			
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,			
RA	Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,			
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevasik E., Vaughan K.,			
RA	Vignati D., Wilcox L., Woldman P., Woldman R., Wilson R.,			
RA	Vaudin M.,			
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	VIII."			
RL	Science 265:2077-2082(1994).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: 000030; AAB68359.1; -			
DR	PIR: S46681; S46681.			
DR	SGD: S0001235; YHR192W.			
KW	Hypothetical protein; Transmembrane.			
FT	TRANSMEM 46 66			
FT	POTENTIAL.			
FT	TRANSMEM 78 98			
FT	POTENTIAL.			
FT	SEQUENCE 278 AA; 32062 MW; D896557EEB805C89 CRC64;			

Query Match 49.5%; Score 45; DB 1; Length 278;  
Best Local Similarity 46.7%; Pred. No. 7.9;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 TVSRVPTWTAAPHGY 15  
|| : || : |||  
Db 62 TVIALSTYTWYEHGY 76

RESULT 2  
YN04\_YEAST

```

ID YN04_YEAST STANDARD: PRT: 350 AA.
AC P53844;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 40.7 kDa protein in PIK1-POL2 intergenic region.
GN YNL264C OR N0815.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SZ88C / FY1679;
RX MEDLINE=96310631; PubMed=8740425;
RA Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes";
RL Yeast 12:505-514(1996).
CC -1- SIMILARITY: BELONGS TO THE SEC14 CYTOSOLIC FACTOR FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X92494; CAA63233.1; -
DR EMBL: Z71540; CAA96171.1; -
DR SGD: S0005208; YNL264C.
DR InterPro: IPR001251; CRAL_TRIO.
DR Pfam: PF00650; CRAL_TRIO.1.
DR SMART: SM00516; SEC14; 1.
DR Hypothetical protein.
KW SEQUENCE 350 AA; 40678 MW; C18F0D225A69DB86 CRC64;
SQ
Query Match 48.4%; Score 44; DB 1; Length 350;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 5 VPTWMAAF 12
Db 247 IPEWMAAF 254
RESULT 3
CHLE_RABIT STANDARD: PRT: 581 AA.
ID CHLE_RABIT STANDARD: PRT: 581 AA.
AC P21927;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cholinesterase precursor (EC 3.1.1.8) (Acylcholine acylhydrolase)
DE (choline esterase II) (Butyrylcholine esterase)
DE (pseudocholinesterase).
DE BCHE.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand;
RX MEDLINE=90326526; PubMed=2374720;
RA Ujilo O., Roudani S., Chaconet A.;
RT "Complete sequence of rabbit butyrylcholinesterase.";
RL Nucleic Acids Res. 18:3990-3990(1990).
RN [2]
RP SEQUENCE OF 75-215 FROM N.A.

```

```

RC TISSUE=Liver;
RA MEDLINE=91201348; PubMed=2016308;
RA Arpagaus M., Chaconet A., Masson P., Newton M., Vaughan T.A.,
RA Bartsels C.F., Nogueira C.P., La Du B.N., Lockridge O.;
RT "Use of the polymerase chain reaction for homology probing of
RT butyrylcholinesterase from several vertebrates.";
RL J. Biol. Chem. 266:6966-6974(1991).
CC -1- CATALYTIC ACTIVITY: An acylcholine + H(2)O = choline + a
CC carboxylic acid anion.
CC SUBUNIT: HOMOTETRAMER. THE TETRAMER IS COMPOSED
CC OF TWO DIMERS. THE TWO SUBUNITS IN A DIMER ARE LINKED BY A
CC DISULFIDE BOND.
CC -1- TISSUE SPECIFICITY: PRESENT IN MOST CELLS EXCEPT ERYTHROCYTES.
CC -1- MISCELLANEOUS: CHOLINESTERASE IS HIGHLY REACTIVE WITH
CC ORGANOPHOSPHATE ESTERS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52090; CAA36308.1; -
DR EMBL: X52091; CAA36308.1; JOINED.
DR EMBL: X52092; CAA36308.1; JOINED.
DR EMBL: M62779; AAA31169.1; -
DR PIR: S10255; S10255.
DR PIR: C39768; C39768.
DR HSP: P21836; IMAH.
DR InterPro: IPR002018; Carboxylesterase_B.
DR InterPro: IPR000997; Cholinesterase.
DR InterPro: IPR000379; Est_1lp_thioest_actsite.
DR Pfam: PF00135; Coesterase.1.
DR PRINTS: PR00878; CHOLINESTERASE.
DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
DR KW Hydrolyase; Serine esterase; Glycoprotein; signal.
FT SIGNAL 1 8
FT CHAIN 9 581
FT ACT_SITE 205 205
FT ACT_SITE 332 332
FT ACT_SITE 445 445
FT ACT_SITE 72 99
FT DISULFID 259 270
FT DISULFID 407 526
FT DISULFID 578 578
FT CARBOHYD 64 64
FT CARBOHYD 113 113
FT CARBOHYD 248 248
FT CARBOHYD 263 263
FT CARBOHYD 348 348
FT CARBOHYD 462 462
FT CARBOHYD 488 488
FT CARBOHYD 492 492
FT CARBOHYD 493 493
FT CARBOHYD 493 493
SQ SEQUENCE 581 AA; 66156 MW; FE8B19E7B32EBOA CRC64;
Query Match 48.4%; Score 44; DB 1; Length 581;
Best Local Similarity 46.7%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;
QY 3 SRVPTWMAAF-AFHGY 15
Db 433 SKLPTWMAAF-AFHGY 447
RESULT 4
CHLE_HUMAN STANDARD: PRT: 602 AA.
ID CHLE_HUMAN STANDARD: PRT: 602 AA.

```



AC p06276; DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 08, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cholinesterase precursor (EC 3.1.1.8) (Acylcholine acylhydrolase)  
DE (choline esterase II) (Butyrylcholine esterase)  
DE (pseudocholinesterase).  
GN BCHE OR CHE1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI.TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90212557; PubMed=232253;  
RA Appagaus M., Kott M., Vatsis K.P., Bartels C.F., la Du B.N.,  
RA Lockridge O.,  
RT "Structure of the gene for human butyrylcholinesterase. Evidence for  
RT a single copy.";  
RL Biochemistry 29:124-131(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87231856; PubMed=3035536;  
RA Prody C.A., Zevin-Sonkin D., Gnatt A., Goldberg O., Soreq H.;  
RT "Isolation and characterization of full-length cDNA clones coding for  
RT cholinesterase from fetal human tissues.";  
RN Proc. Natl. Acad. Sci. U.S.A. 84:3555-3559(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Brain;  
RA MEDLINE=88016155; PubMed=3477799;  
RA McIernan C., Adkins S., Chaconnet A., Vaughan T.A., Bartels C.F.,  
RA Kott M., Rosenberry T.L., la Du B.N., Lockridge O.;  
RT "Brain cDNA clone for human cholinesterase.";  
RN Proc. Natl. Acad. Sci. U.S.A. 84:6682-6686(1987).  
RN [4]  
RP SEQUENCE OF 29-602.  
RX TISSUE=Plasma;  
RA MEDLINE=87109144; PubMed=3542989;  
RA Lockridge O., Bartels C.F., Vaughan T.A., Wong C.K., Norton S.E.,  
RA Johnson L.L.;  
RT "Complete amino acid sequence of human serum cholinesterase.";  
RN J. Biol. Chem. 262:549-557(1987).  
RN [5]  
RP DISULFIDE BONDS.  
RX MEDLINE=88007487; PubMed=3115973;  
RA Lockridge O., Adkins S., la Du B.N.;  
RT "Location of disulfide bonds within the sequence of human serum  
RT cholinesterase.";  
RN J. Biol. Chem. 262:12945-12952(1987).  
RN [6]  
RP REVIEW.  
RX MEDLINE=89149758; PubMed=3067729;  
RA Lockridge O.;  
RT "Structure of human serum cholinesterase.";  
RN Bioessays 9:125-128(1988).  
RN [7]  
RP VARIANT ATYPICAL GUY-98.  
RX MEDLINE=89128896; PubMed=2915989;  
RA McGuire M.C., Nogueira C.P., Bartels C.F., Lightstone H., Hajta A.,  
RA van der Spek A.F.L., Lockridge O., la Du B.N.;  
RT "Identification of the structural mutation responsible for the  
RT dibucaine resistant (atypical) variant form of human serum  
RT cholinesterase.";  
RN Proc. Natl. Acad. Sci. U.S.A. 86:953-957(1989).  
RN [8]  
RP VARIANT ILE-358.  
RX MEDLINE=96287386; PubMed=8680411;  
RA Iida S., Kinoshita M., Fujii H., Moriyama Y., Nakamura Y., Yura N.,  
RA Moriwaki K.;  
RT "Mutations of human butyrylcholinesterase gene in a family with  
RT hypocholesterolemia.";

Rt	Hm.	Mutalt	6:349-351(1995).
CC	-1-	CATALYTIC ACTIVITY:	An acylcholine + H(2)O = choline + a
CC	-1-	carboxylic acid anion.	
CC	-1-	SUBUNIT:	HOMOTETRAMER. THE TETRAMER IS COMPOSED
CC	-1-	OF TWO DIMERS.	THE TWO SUBUNITS IN A DIMER ARE LINKED BY A
CC	-1-	DISULFIDE BOND.	
CC	-1-	TISSUE SPECIFICITY:	PRESSENT IN MOST CELLS EXCEPT ERYTHROCYTES.
CC	-1-	DISEASE:	MUTANT ALLELES OF CHE1 ARE RESPONSIBLE FOR
CC	-1-	HYPCHOLINESTERASEMIA RESULTING	IN SUXAMETHONIUM SENSITIVITY.
CC	-1-	HOMOZYGOUS PERSONS SUSTAIN	PROLONGED APNEA AFTER ADMINISTRATION
CC	-1-	OF THE MUSCLE RELAXANT	SUXAMETHONIUM IN CONNECTION WITH SURGICAL
CC	-1-	ANESTHESIA.	
CC	-1-	MISCELLANEOUS:	CHOLINESTERASE IS HIGHLY REACTIVE WITH
CC	-1-	ORGANOPHOSPHATE ESTERS.	
CC	-1-	SIMILARITY:	BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
CC	-1-	This SWISS-PROT entry is copyright.	It is produced through a collaboration
CC	-1-	between the Swiss Institute of Bioinformatics	and the EMBL Outstation -
CC	-1-	the European Bioinformatics Institute.	There are no restrictions on its
CC	-1-	use by non-profit institutions as long as	its content is in no way used for
CC	-1-	modified and this statement is not removed.	usage by and for commercial
CC	-1-	entities requires a license agreement (see	http://www.isb-sib.ch/announce/
CC	-1-	or send an email to licensese@isb-sib.ch).	
DR	EMBL;	M32381;	AAA99296.1; -
DR	EMBL;	M32389;	AAA99296.1; JOINED.
DR	EMBL;	M32390;	AAA99296.1; JOINED.
DR	EMBL;	M16541;	AAA98113.1; -
DR	EMBL;	M16474;	AAA52013.1; -
DR	PIR;	A00772;	ACHU.
DR	PIR;	A26613;	A26613.
DR	PIR;	A33769;	A33769.
DR	PIR;	A33887;	A33887.
DR	HSSP;	P21836;	IAMA.
DR	MM;	177400;	-
DR	InterPro;	IPR002018;	Carboxylesterase_B.
DR	InterPro;	IPR000997;	Cholinesterase.
DR	InterPro;	IPR000379;	Esl_1hp_thioest_actsite.
DR	Pfam;	PF00133;	Coesterase; 1.
DR	PRINTS;	PR00878;	Cholinesterase.
DR	PROSITE;	PS00122;	CARBOXYLESTERASE_B.1; 1.
DR	PROSITE;	PS00941;	CARBOXYLESTERASE_B.2; 1.
Kw	Hydrolase;	Serine esterase;	Glycoprotein; Signal; Disease mutation;
Kw	Polymorphism.		
FT	CHAIN	1	28
FT	ACT_SITE	229	602
FT	ACT_SITE	226	226
FT	ACT_SITE	353	353
FT	ACT_SITE	466	466
FT	DISULFID	93	120
FT	DISULFID	280	291
FT	DISULFID	428	547
FT	DISULFID	599	599
FT	CARBOHYD	45	45
FT	CARBOHYD	85	85
FT	CARBOHYD	134	134
FT	CARBOHYD	269	269
FT	CARBOHYD	284	284
FT	CARBOHYD	369	369
FT	CARBOHYD	483	483
FT	CARBOHYD	509	509
FT	CARBOHYD	514	514
FT	VARIANT	98	98
FT	VARIANT	271	271
FT	VARIANT	358	358
FT	VARIANT	418	418
FT	VARIANT	567	567

```

FT SEQUENCE 602 AA; 68418 MW; C9836409D9057E27 CRC64;
SQ

Query Match 48.4%; Score 44; DB 1; Length 602;
Best Local Similarity 46.7%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 3 SRVPTAM--AFHGY 15
   |||
Db 454 SKLPWPMGVMHGY 468

RESULT 5
CHIE_MOUSE STANDARD; PRT; 603 AA.
ID 003311;
AC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cholinesterase precursor (EC 3.1.1.8) (Acylcholine acylhydrolase)
DE (Choline esterase II) (Butyrylcholine esterase)
DE (pseudocholinesterase).
GN BChE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90380429; PubMed=2400605;
RA Rachinsky T.L., Camp S., Li Y., Ekstrom T.J., Newton M., Taylor P.;
RT "Molecular cloning of mouse acetylcholinesterase: tissue distribution
RT of alternatively spliced mRNA species.";
RN Neuron 5:317-327(1990).
RL [2]
RP SEQUENCE OF 97-237 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91201348; PubMed=2016308;
RA Appagou M., Chatonnet A., Masson P., Newton M., Vaughan T.A.,
RA Bartels C.F., Nogueira C.P., La Du B.N., Lockridge O.;
RT "Use of the polymerase chain reaction for homology probing of
RT butyrylcholinesterase from several vertebrates.";
RL J. Biol. Chem. 266:6966-6974(1991).
CC -1- CATALYTIC ACTIVITY: An acylcholine + H(2)O = choline + a
CC carboxylic acid anion.
CC -1- SUBUNIT: HOMOTETRAMER. THE TETRAMER IS COMPOSED
CC OF TWO DIMERS. THE TWO SUBUNITS IN A DIMER ARE LINKED BY A
CC DISULFIDE BOND.
CC -1- TISSUE SPECIFICITY: PRESENT IN MOST CELLS (EXCEPT ERYTHROCYTES).
CC -1- MISCELLANEOUS: CHOLINESTERASE IS HIGHLY REACTIVE WITH
CC ORGANOPHOSPHATE ESTERS.
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M99492; AAA37328.1; -
DR PIR, A39768; A39768.
DR HSSP, P21836; 1MAH.
DR MGD, MGI:894278; Bche.
DR InterPro: IPR002018; Carboxylesterase_B.
DR InterPro: IPR000997; Cholinesterase.
DR InterPro: IPR000379; Est_11p_choest_actsite.
DR Pfam: PF00133; Coesterase_1.
DR PRINTS: PR00878; CHOLINESTRASE.
DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.

```

```

KM Hydrolyase; Serine esterase; Glycoprotein; Signal.
FT SIGNAL 1 29
FT CHAIN 30 603
FT ACT_SITE 227 227 CHOLINESTERASE.
FT ACT_SITE 354 354 BY SIMILARITY.
FT ACT_SITE 467 467 BY SIMILARITY.
FT ACT_SITE 467 467 BY SIMILARITY.
FT DISULFID 94 121 BY SIMILARITY.
FT DISULFID 281 292 BY SIMILARITY.
FT DISULFID 429 548 BY SIMILARITY.
FT DISULFID 600 600 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 510 510 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 129 129 R -> P (IN REF. 2).
SQ SEQUENCE 603 AA; 68521 MW; 719B1B220D1E5367 CRC64;

```

```

Query Match 48.4%; Score 44; DB 1; Length 603;
Best Local Similarity 46.7%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

```

```

OY 3 SRVPTAM--AFHGY 15
   |||
Db 455 SKLPWPMGVMHGY 469

RESULT 6
VPL_BPP2 STANDARD; PRT; 176 AA.
ID VPL_BPP2
AC P26701;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Tail protein I (GPI).
GN I.
OS Bacteriophage P2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC P2-like viruses.
OX NCBI_TaxID=10679;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92165720; PubMed=1531648;
RA Haggard-Ljungquist E., Halling C., Calendar R.;
RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence
RT for horizontal transfer of tail fiber genes among unrelated
RT bacteriophages.";
RL J. Bacteriol. 174:1462-1477(1992).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96036485; PubMed=7483254;
RA Haggard-Ljungquist E., Jacobsen E., Rishoyd S., Six E.W., Nilsen O.,
RA Sunshine M.G., Lindqvist B.H., Kim K.-J., Barreiro V., Koonin E.V.,
RA Calendar R.;
RT "Bacteriophage P2: genes involved in baseplate assembly.";
RL Virology 213:109-121(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF063097; AAD03285.1; -
DR Structural Protein.
SQ SEQUENCE 176 AA; 19584 MW; EADA32BCAFCECF1E CRC64;

```

Query Match 47.3%; Score 43; DB 1; Length 176;  
 Best Local Similarity 75.0%; Pred. No. 10;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 VPMWMAF 12  
 : || || || ||  
 DB 46 LPMWMAF 53

## RESULT 7

YEO6\_YEAST STANDARD; PRT; 302 AA.  
 AC P40049;

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 33.5 kDa protein in PTP3-SER3 intergenic region  
 DE precursor.  
 GN YER076C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;

[1]

SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,  
 RA Aviles E., Berne A., Brennan T., Carpenter J., Chen E., Cherry J.M.,  
 RA Chung E., Duncan M., Guzman E., Hartzell D., Hunkle-Smith S.,  
 RA Hyman R., Kayser A., Komp C., Laskari D., Lew H., Lin D.,  
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,  
 RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

-1- SIMILARITY: TO YEAST KILLER TOXIN KHR.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
 CC EMBL: U18839; AAB64631.1; -

DR SCD: S0000878; YER076C.  
 KW Hypothetical protein; Glycoprotein; Signal.

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

FT SIGNAL

Query Match 47.3%; Score 43; DB 1; Length 302;  
 Best Local Similarity 75.0%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 7 WPMWMAF 14  
 : || || || ||  
 DB 180 WPMWMAF 187

## RESULT 8

YXN1\_YEAST STANDARD; PRT; 351 AA.  
 AC P53860;

DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 40.7 kDa protein in CSL4-URE2 intergenic region.  
 GN YNL231C OR N1158.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;

[1]

SEQUENCE FROM N.A.  
 RX MEDLINE=97051596; PubMed=8896273;  
 RA Pandolfo D., de Antoni A., Lanfranchi G., Valle G.;

RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open  
 RT reading frames including a novel gene encoding a globin-like  
 RT domain.";  
 RL Yeast 12:1071-1076 (1996).

-1- SIMILARITY: BELONGS TO THE SEC14 CYTOSOLIC FACTOR FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
 CC EMBL: Z69381; CAA93367.1; -

DR EMBL: Z71507; CAA96136.1; -

DR SCD: S0005175; YNL231C

DR InterPro: IPR001251; CRAL\_TRIO.

DR Pfam: PF00650; CRAL\_TRIO; 1.

DR SMART: SM00516; SEC14; 1.

DR Hypothetical protein.

SQ SEQUENCE 351 AA; 40714 MW; 24C5B3262016F037 CRC64;

Query Match 47.3%; Score 43; DB 1; Length 351;  
 Best Local Similarity 45.5%; Pred. No. 19;  
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 VSRVPMWMAF 12  
 : : || || || ||  
 DB 242 LTRVPMWMAF 252

## RESULT 9

GUN3\_FIBSU STANDARD; PRT; 658 AA.  
 AC P14250;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Endoglucanase 3 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)

DE (Cellulase) (EG3).

GN CEL-3.

OS Fibrobacter succinogenes (Bacteroides succinogenes).

OC Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;

OC Fibrobacter.

OX NCBI\_TaxID=833;

[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 266-287.

RP STRAIN=585;

RX MEDLINE=9008798; PubMed=2676979;

RA McGavin M.J., Forsberg C.W., Crosby B., Bell A.W., Dignard D.,  
 RA Thomas D.V.;

RT "Structure of the cel-3 gene from Fibrobacter succinogenes S85 and  
 RT characteristics of the encoded gene product, endoglucanase 3.";

RL J. Bacteriol. 171:5587-5595 (1989).

-1- FUNCTION: EXHIBITS BOTH ENDOGLUCANASE AND CELLOBIOSIDASE

CC ACTIVITIES.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

CC linkages in cellulose.

CC -1- SUBUNIT: MONOMER.

CC -1- PTM: MAY BE A LIPOPROTEIN AND MAY BE GLYCOSYLATED.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL

```

CC -----
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29047; AAA24893.1; -.
CC PIR; A33598; A33598.
CC HSSP; P07985; 1CEN.
CC -----
CC InterPro; IPR001547; Glyco_hydro.F5.
CC Pfam; PF00150; cellulase; 1.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Lipoprotein;
KM Signal.
FT SIGNAL. 1 23 POTENTIAL.
FT PROPEP 24 265 ENDOGLUCANASE 3.
FT CHAIN 266 658 N-ACYL DIGLYCERIDE (POTENTIAL).
FT LIPID 24 24 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 448 448 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 597 597
SQ SEQUENCE 658 AA; 73424 MW; 1C96E64C3F7109A6 CRC64;

Query Match 47.3%; Score 43; DB 1; Length 658;
Best Local Similarity 45.5%; Pred. No. 33;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 RVPWTAMAFHG 14
:|||||:
Db 625 QIPWTHMGYTG 635

RESULT 10
RPOW_SCHPO STANDARD; PRT; 1120 AA.
AC 013993;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNA-directed RNA polymerase, mitochondrial precursor (EC 2.7.7.6).
GN SPAC26H5.12.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RL Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC - CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC - SUBCELLULAR LOCATION: Mitochondrial.
CC - SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; Z99126; CAB16197.1; -.
DR HSSP; P00573; IARO.
DR InterPro; IPR002092; RNA_pol_phage.
DR Pfam; PF00940; RNA_pol.1.
DR PROSITE; PS00900; RNA_POL_PHAGE_1; 1.
DR PROSITE; PS00489; RNA_POL_PHAGE_2; 1.
CC Translase; DNA-directed RNA polymerase; Transcription;
KM Mitochondrion; Transl peptide.
FT TRANSIT 1 2 MITOCHONDRION.
FT CHAIN 787 787 BY SIMILARITY.
FT ACT_SITE 787 787 BY SIMILARITY.
FT ACT_SITE 856 856 BY SIMILARITY.
FT ACT_SITE 1027 1027 BY SIMILARITY.
SQ SEQUENCE 1120 AA; 127268 MW; 22D206A1F2831A23 CRC64;

Query Match 47.3%; Score 43; DB 1; Length 1120;
Best Local Similarity 50.0%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 VSRVPTAMAFHG 15
:|||||:
Db 519 VTRPWTSMIDGY 532

RESULT 11
YUG8_YEAST STANDARD; PRT; 299 AA.
ID YUG8_YEAST
AC P40363;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 33.9 kDa esterase in SMC3-MRPL8 intergenic region
DE (EC 3.1.1.*).
GN YU068C OR J1102 OR HRE299.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=S288C;
RX MEDLINE=95282514; Pubmed=7762302;
RA Vandenbol M., Durand P., Dion C., Portetelle D., Hilger F.;
RT "Sequence of a 17.1 kb DNA fragment from chromosome X of
RT Saccharomyces cerevisiae includes the mitochondrial ribosomal protein
RT L8."
RT Yeast 11:57-60(1995).
RN [2]
RP SEQUENCE OF 15-299 FROM N.A.
RC STRAIN=S288C;
RA sor F.J.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: STRONG. TO HUMAN ESTERASE D.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z34288; CA84054.1; -.
DR EMBL; Z49343; CA89359.1; -.
DR EMBL; X88851; CA61307.1; -.
DR PIR; S47122; S47122.
DR SGD; S0003604; YU068C.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR InterPro; IPR000801; Esterase_put.
DR Pfam; PF00756; Esterase; 1.
KM Hypothetical protein; Hydrolase.
FT ACT_SITE 161 161 BY SIMILARITY.

```

```

SQ SEQUENCE 299 AA; 33934 MW; BFA5DC02729E9AEF CRC64;

Query Match 46.2%; Score 42; DB 1; Length 299;
Best Local Similarity 61.5%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 SRVPTAMAFHG 15
   | | | | | | |
Db 193 SNVPMGQKAFKGY 205

RESULT 12
YA35_MYCPN
ID YA35_MYCPN STANDARD; PRT; 666 AA.
AC P75079;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN035 (B01_orf666).
GN MPN035 OR MP119.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreid R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE MG032 / MG096 / MG288 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000014; AAB95767.1; -
DR InterPro: IPR004306; DUF237.
DR InterPro: IPR004319; DUF240.
DR Pfam: PF03072; DUF237; 1.
DR Pfam: PF03086; DUF240; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 666 AA; 76898 MW; DF096AA4B475402D CRC64;

Query Match 46.2%; Score 42; DB 1; Length 666;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 VPTAMAFHG 14
   | | | | | |
Db 557 VAWMAAFKG 566

RESULT 13
CTPB_MYCLE
ID CTPB_MYCLE STANDARD; PRT; 750 AA.
AC P46840;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cation-transporting P-type ATPase B (EC 3.6.3.-).
GN CTPB OR M2200.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

FT CONFLICT 544 548 EGRSR -> RRRIT (IN REF. 1).  
 SQ SEQUENCE 750 AA; 78100 MW; ID58477/D4A69B00D CRC64;  
 Query Match 46.2%; Score 42; DB 1; Length 750;  
 Best Local Similarity 66.7%; Pred. No. 52;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 5 VPMWMAFH 13  
 1 1 1 1 1  
 Db 144 VPMWMAFH 152  
 RESULT 14  
 CTPA\_MYCTU STANDARD; PRT; 752 AA.  
 ID CTPA\_MYCTU  
 AC Q10877; P77905;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Cation-transporting P-type ATPase B (EC 3.6.3.-).  
 GN CTPB OR RV0103C OR MT0112 OR MTCY251.22C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NC NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA Aranooff D.D., Krishna S.K., Griffin G.E., Mangan J.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Stulson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RT Nature 393:537-544(1998).  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RP CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (E1-E2 ATPASES). SUBFAMILY IB.  
 CC -1- SIMILARITY: CONTAINS 1 HMA DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y07638; CAA68915.1; -;  
 CC DR EMBL: 274410; CAA98939.1; -;

DR EMBL: AE006923; AAK44334.1; -;  
 DR TIGR: MT0112; -;  
 DR TubercuList; RV0103C; -;  
 DR InterPro; IPR000579; Cat\_P\_ATPaseA.  
 DR InterPro; IPR001757; E1-E2\_ATPase.  
 DR InterPro; IPR001934; HMA.  
 DR InterPro; IPR001454; Hydrolase.  
 DR Pfam; PF00122; E1-E2\_ATPase; 1.  
 DR Pfam; PF00403; HMA; 1.  
 DR Pfam; PF00702; Hydrolase; 1.  
 DR PRINTS: PR00940; CATPATPASEA.  
 DR PROSITE: PS00154; ATPASE\_E1\_E2; 1.  
 DR PROSITE; PS01047; HMA\_1; 1.  
 DR PROSITE; PS50846; HMA\_2; 1.  
 DR Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
 KW Complete proteome.  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 132 152 POTENTIAL.  
 FT TRANSMEM 167 187 POTENTIAL.  
 FT TRANSMEM 201 221 POTENTIAL.  
 FT TRANSMEM 361 381 POTENTIAL.  
 FT TRANSMEM 390 410 POTENTIAL.  
 FT TRANSMEM 491 511 POTENTIAL.  
 FT TRANSMEM 714 734 POTENTIAL.  
 FT DOMAIN 16 79 HMA.  
 FT MOD\_RES 446 446 PHOSPHORYLATION (BY SIMILARITY).  
 FT METAL 26 446 BY SIMILARITY.  
 FT METAL 29 29 BY SIMILARITY.  
 FT CONFLICT 22 22 L -> S (IN REF. 3).  
 SQ SEQUENCE 752 AA; 77509 MW; 21563FEA84407505 CRC64;  
 Query Match 46.2%; Score 42; DB 1; Length 752;  
 Best Local Similarity 66.7%; Pred. No. 52;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 5 VPMWMAFH 13  
 1 1 1 1 1  
 Db 145 VPMWMAFH 153  
 RESULT 15  
 CTPA\_MYCTU STANDARD; PRT; 761 AA.  
 ID CTPA\_MYCTU  
 AC Q10876;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Cation-transporting P-type ATPase A (EC 3.6.3.-).  
 GN CTPA OR RV0052 OR MT0101 OR MTCY251.11.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NC NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Stulson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RT Nature 393:537-544(1998).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -I- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.  
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -I- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 (EI-E2 ATPASES). SUBFAMILY IB.  
 CC -I- SIMILARITY: CONTAINS 1 HMA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z74410; CAA98928.1; -  
 DR EMBL: AE006921; AAK4324.1; -  
 DR TIGR: MT0101; -  
 DR TuberculList; RV0092; -  
 DR InterPro: IPR000579; Cat\_P\_ATPase.  
 DR InterPro: IPR001757; EI-E2\_ATPase.  
 DR InterPro: IPR001934; HMA.  
 DR InterPro: IPR001454; Hydrolyase.  
 DR Pfam; PF00122; EI-E2\_ATPase; 1.  
 DR Pfam; PF00403; HMA; 1.  
 DR PRINTS: PR00940; CATPATPASEA.  
 DR PROSITE: PS00154; ATPASE\_EI\_E2; 1.  
 DR PROSITE: PS01047; HMA\_1; 1.  
 DR PROSITE: PS0846; HMA\_2; 1.  
 DR Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
 KW Complete proteome.  
 FT TRANSMEM 102 122 POTENTIAL.  
 FT TRANSMEM 129 149 POTENTIAL.  
 FT TRANSMEM 164 184 POTENTIAL.  
 FT TRANSMEM 199 219 POTENTIAL.  
 FT TRANSMEM 361 381 POTENTIAL.  
 FT TRANSMEM 387 407 POTENTIAL.  
 FT TRANSMEM 548 568 POTENTIAL.  
 FT TRANSMEM 650 670 POTENTIAL.  
 FT TRANSMEM 703 723 POTENTIAL.  
 FT DOMAIN 13 79 HMA.  
 FT MOD\_RES 443 443 PHOSPHORYLATION (BY SIMILARITY).  
 FT METAL 25 25 BY SIMILARITY.  
 FT METAL 28 28 BY SIMILARITY.  
 SQ SEQUENCE 761 AA: 78850 MW: D397901EC5C513BE CRC64:

Query Match 46.2%; Score 42; DB 1; Length 761;  
 Best Local Similarity 66.7%; Pred. No. 53;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 5 VPWTAMAFH 13  
 | | | | |  
 Db 142 YTWAMAFH 150

Search completed: August 15, 2002, 11:47:41  
 Job time: 517 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:38:29 ; Search time 58.4 Seconds  
(without alignments)  
44.434 Million cell updates/sec

Title: US-09-613-092a-5

Perfect score: 91

Sequence: 1 TVSRVPMWTAMAFHCY 15

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPRMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	49.5	297	10	Q9FTE7
2	45	49.5	413	2	Q9K3G1
3	45	49.5	454	10	Q9X1S9
4	45	49.5	537	10	Q9SD75
5	45	49.5	890	3	Q9HGE4
6	44	48.4	349	6	Q9GKJ6
7	44	48.4	597	11	Q9JRC1
8	44	48.4	603	13	Q9OZK8
9	43.5	47.8	310	10	Q8Z590
10	43.5	47.8	457	10	Q9M126
11	43	47.3	176	9	Q37841
12	43	47.3	344	10	Q9LSP1
13	43	47.3	392	16	Q92MA2
14	43	47.3	433	16	Q9A0U1
15	43	47.3	441	16	Q9HYE2
16	43	47.3	505	16	Q9JYF5

17	43	47.3	505	16	Q9JTE0	Q9JTE0 neisseria m
18	43	47.3	572	2	P72076	P72076 neisseria g
19	43	47.3	602	6	O62760	O62760 felis silve
20	43	47.3	602	6	O62761	O62761 panthera ti
21	43	47.3	657	2	O59446	O59446 fibrobacter
22	43	47.3	668	2	O9ALX8	O9ALX8 burkholderi
23	43	47.3	669	2	O59445	O59445 fibrobacter
24	43	47.3	722	10	O04514	O04514 arabisdops
25	43	47.3	938	10	O9AV04	O9AV04 oryza sativ
26	42	46.2	81	5	Q9VBC3	Q9VBC3 dirosophila
27	42	46.2	141	11	Q9CZAL	Q9CZAL mus musculu
28	42	46.2	142	10	Q9SUN7	Q9SUN7 arabisdops
29	42	46.2	247	17	O27503	O27503 methanot
30	42	46.2	297	16	Q9KDS0	Q9KDS0 bacillus ha
31	42	46.2	317	16	O9XES0	O9XES0 bacillus ha
32	42	46.2	548	4	O14818	O14818 homo sapien
33	42	46.2	602	5	O76999	O76999 brachiosto
34	42	46.2	623	4	O14819	O14819 homo sapien
35	42	46.2	650	10	O23349	O23349 arabisdops
36	42	46.2	668	10	O48552	O48552 arabisdops
37	42	46.2	884	12	O98XW8	O98XW8 elmeria bru
38	41	45.1	130	2	O9X4J8	O9X4J8 salmonella
39	41	45.1	141	2	O9WTB5	O9WTB5 escherichia
40	41	45.1	164	10	O9FSI8	O9FSI8 hordeum vul
41	41	45.1	174	2	O9K414	O9K414 streptomyce
42	41	45.1	232	2	O9AHG5	O9AHG5 comanona t
43	41	45.1	232	2	P71191	P71191 enterobacte
44	41	45.1	305	2	O9FI32	O9FI32 pseudomonas
45	41	45.1	385	16	O98HGI	O98HGI rhizobium 1

## ALIGNMENTS

RESULT 1  
Q9FTE7 PRELIMINARY; PRT; 297 AA.  
AC Q9FTE7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE P0698A04.4 PROTEIN (P0494A10.20 PROTEIN).  
GN P0698A04.4 OR P0494A10.20.  
OS Oryza sativa (Rice).  
OC Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsunoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(Ga3) genomic DNA, chromosome 1, PAC  
clone:P0698A04.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsunoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(Ga3) genomic DNA, chromosome 1, PAC  
clone:P0494A10.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002868; BAB17143.1; -;  
DR EMBL; AP002541; BAB55480.1; -;  
DR InterPro; IPR004277; PSS.  
DR Pfam; PF03034; PSS; 1.  
SQ SEQUENCE 297 AA; 34429 MW; E1730BFC71208E09 CRC64;

Query Match 49.5%; Score 45; DB 10; Length 297;  
Best Local Similarity 85.7%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PWTAMAF 12  
 |||||:  
 Db 32 PWTAMAF 38

## RESULT 2

Q9K3G1 PRELIMINARY; PRT; 413 AA.  
 AC Q9K3G1;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN.  
 GN 25CG18.27C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 NX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL390188; CAB99160.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 413 AA; 44536 MW; 2C841BDA772FCH3D CRC64;

Query Match 49.5%; Score 45; DB 2; Length 413;  
 Best Local Similarity 50.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TVSRVPTWAMAFHG 14  
 | | | | | | | | | |  
 Db 113 TDDRPPWDVWLVRHC 126

RESULT 3  
 Q9X159  
 ID Q9X159 PRELIMINARY; PRT; 454 AA.  
 AC Q9X159;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE F9L1.4 PROTEIN.  
 GN F9L1.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
 RA Lee J., Liu A., Li J., Kremenetskaia I., Luros J., Gonzalez A.,  
 RA Alatafi H., Araujo R., Brooks S., Buehler E., Chao Q., Conn L.,  
 RA Conway A.B., Dunn P., Hansen N., Hutzler L., Khan S., Kim C., Palm C.,  
 RA Rowley D., Shinn P., Walker M., Davis R.W., Becker J.R.,

RA Federspiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F9L1 sequence."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC007591; AAD39639.1; -;  
 DR InterPro: IPR004277; PSS.  
 DR Pfam: PF03034; PSS; 1.  
 SQ SEQUENCE 454 AA; 52610 MW; 1EDC8AA5A4188A44 CRC64;

Query Match 49.5%; Score 45; DB 10; Length 454;  
 Best Local Similarity 85.7%; Pred. No. 41;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PWTAMAF 12  
 |||||:  
 Db 32 PWTAMAF 38

## RESULT 4

Q9SD75 PRELIMINARY; PRT; 537 AA.  
 AC Q9SD75;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 58.5 KDA PROTEIN.  
 GN F1312.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choiane N., Robert C., Brotlier P., Mincker P., Catrolicco L.,  
 RA Attiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X., Queller F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC - SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 DR EMBL: AL133292; CAB61944.1; -;  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr; 1.  
 KW Hypothetical protein; Transmembrane.  
 SQ SEQUENCE 537 AA; 58503 MW; E4D85E54B9429A67 CRC64;

Query Match 49.5%; Score 45; DB 10; Length 537;  
 Best Local Similarity 61.5%; Pred. No. 48;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 SRVPWTAMAFHG 15  
 | | | | | | | | | |  
 Db 363 SAVPWSMAAFMG 375

## RESULT 5

Q9HGE4 PRELIMINARY; PRT; 890 AA.  
 AC Q9HGE4;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE CHITIN SYNTHASE 1.  
 OS Arthropoda; Insecta; Arthropoda; Insecta; Arthropoda; Insecta;  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Onygenales; Arthrodermataceae; Arthroderma.  
 NX NCBI\_TaxID=63400;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-VUT-77011;  
RA Kano R.;  
RT "Arhoderma benhamiae gene for chitin synthase 1, complete."  
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB050933; BAB17766.1; -;  
DR InterPro: IPR002923; Chitin\_synth.  
DR InterPro: IPR001173; Glycos\_transf\_2.  
DR Pfam: PF01644; Chitin\_synth.1.  
DR ProDom: PD002998; Chitin\_synth.1.  
SQ SEQUENCE 890 AA: 101598 MW: 547C38823E64B565 CRC64;

Query Match 49.5%; Score 45; DB 3; Length 890;  
Best Local Similarity 46.7%; Pred. No. 78;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 TSVRPWTAW-AFHGY 15  
DB 154 SISRIPLTRMTODGY 168

RESULT 6  
O9GKJ6 PRELIMINARY; PRT; 349 AA.

AC 09GKJ6; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
DE BUTYRYLCHOLINESTERASE (FRAGMENT).  
GN BChE.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Van Poucke M., Yerle M., Tuglie C., Chardon P., Van Zeveren A.,  
RA Peelam L.J.;  
RT "Integration of porcine chromosome 13 maps."  
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR EMBL: AF222914; AAG41127.1; -;  
DR HSSP: P21836; IMAA.  
DR InterPro: IPR002018; Carboxylesterase\_B.  
DR InterPro: IPR000997; Cholinesterase.  
DR InterPro: IPR000379; Est\_lip\_thioest\_acsite.  
DR Pfam: PF00135; Coesterase; 1.  
DR PRINTS: PR00878; CHOLINESTRASE.  
DR PROSITE: PS00122; CARBOXYLESTERASE\_B\_1; 1.  
KW Hydrolyase.  
FT NON\_TER 349  
FT NON\_TER 349  
SQ SEQUENCE 349 AA: 39061 MW: D66354B14725BE58 CRC64;

Query Match 48.4%; Score 44; DB 6; Length 349;  
Best Local Similarity 46.7%; Pred. No. 45;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 3 SRVPWTAW-AFHGY 15  
DB 314 SKLPPEWGMGVHGY 328

RESULT 7  
O9JKC1 PRELIMINARY; PRT; 597 AA.

AC 09JKC1; 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
DE BUTYRYLCHOLINESTERASE.

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEART;  
RA Li B., Stribley J., Tieu A., Xie W., Schopfer L.M., Hammond P.,  
RA Brindoin S., Hinrichs S.H., Lockridge O.;  
RT "Abundant Tissue Butyrylcholinesterase and its Possible Function in  
the Acetylcholinesterase Knockout Mouse."  
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR EMBL: AF244349; AAF44713.1; -;  
DR HSSP: P21836; IMAA.  
DR InterPro: IPR002018; Carboxylesterase\_B.  
DR InterPro: IPR000997; Cholinesterase.  
DR InterPro: IPR000379; Est\_lip\_thioest\_acsite.  
DR Pfam: PF00135; Coesterase; 1.  
DR PRINTS: PR00878; CHOLINESTRASE.  
DR PROSITE: PS00122; CARBOXYLESTERASE\_B\_1; 1.  
DR PROSITE: PS00941; CARBOXYLESTERASE\_B\_2; 1.  
KW Hydrolyase.  
SQ SEQUENCE 597 AA: 67776 MW: 771204D166CTEAC CRC64;

Query Match 48.4%; Score 44; DB 11; Length 597;  
Best Local Similarity 46.7%; Pred. No. 76;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 3 SRVPWTAW-AFHGY 15  
DB 449 SKLPPEWGMGVHGY 463

RESULT 8  
O90ZK8 PRELIMINARY; PRT; 603 AA.

AC 090ZK8; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
DE BUTYRYLCHOLINESTERASE PRECURSOR (EC 3.1.1.8).  
GN BChE.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Geisler K., Chatonnet A., Layer P.G.;  
RT "Chicken Butyrylcholinesterase."  
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ306928; CAC37792.1; -;  
KW Signal: Hydrolyase.  
FT SIGNAL 1  
FT CHAIN 30 603 POTENTIAL.  
FT CHAIN 30 603 BUTYRYLCHOLINESTERASE.  
SQ SEQUENCE 603 AA: 68480 MW: A350FDF68574ADF CRC64;

Query Match 48.4%; Score 44; DB 13; Length 603;  
Best Local Similarity 46.7%; Pred. No. 76;  
Matches 7; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

OY 3 SRVPWTAW-AFHGY 15  
DB 454 SKLPPEWGMGVHGY 468

RESULT 9  
O82590 PRELIMINARY; PRT; 310 AA.

AC 082590;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE F1104.5 PROTEIN.  
 GN F1104.5  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA WashU;  
 RT "The A. thaliana Genome Sequencing Project."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Abu-Threideh J., Stoneking T., Langston Y., Trevasakis E.;  
 RT "The sequence of A. thaliana F1104."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Waterston R.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF096370; AAC62781.1; -;  
 DR InterPro: IPR003441; NAM.  
 DR Pfam: PF02365; NAM.1  
 SQ SEQUENCE 310 AA; 35370 MW; 4CE95002CD638F32 CRC64;

Query Match 47.8%; Score 43.5; DB 10; Length 310;  
 Best Local Similarity 53.3%; Pred. No. 48;  
 Matches 8; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 4 RVP---WTANAFHCY 15  
 ||| || | |  
 Db 95 RVPKGVWTPWMVMEY 109

RESULT 10  
 Q9M126 PRELIMINARY; PRT; 457 AA.  
 ID Q9M126  
 AC Q9M126;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE PUTATIVE NAM-LIKE PROTEIN.  
 GN AT4G01550.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL161492; CAB77725.1; -;  
 DR InterPro: IPR003441; NAM.  
 DR Pfam: PF02365; NAM.1  
 SQ SEQUENCE 457 AA; 51816 MW; 9CBBCE615F0538B3 CRC64;

Query Match 47.8%; Score 43.5; DB 10; Length 457;

Best Local Similarity 53.3%; Pred. No. 69;  
 Matches 8; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 4 RVP---WTANAFHCY 15  
 ||| || | |  
 Db 119 RVPKGVWTPWMVMEY 133

RESULT 11  
 Q37841 PRELIMINARY; PRT; 176 AA.  
 ID Q37841  
 AC Q37841;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE ORF38.  
 OS Bacteriophage 186.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC P2-like viruses;  
 OX [1]  
 RN NCBI\_TaxID=29252;  
 RP SEQUENCE OF 174-176 FROM N.A.  
 RX MEDLINE=95407087; PubMed=7676622;  
 RA Xue Q., Egan J.B.;  
 RT "DNA sequence of tail fiber genes of coliphage 186 and evidence for a  
 common ancestor shared by dsDNA phage fiber genes."  
 RL Virology 212:128-133(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Xue Q.;  
 RT "Studies on the tail region of the temperate coliphage 186 genome."  
 RL Thesis (1993), University of Adelaide.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98371265; PubMed=9705261;  
 RA Portelli R., Dodd I.B., Xue Q., Egan J.B.;  
 RT "The late-expressed region of the temperate coliphage 186 genome."  
 RL Virology 248:117-130(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Dodd I.B., Egan J.B.;  
 RT "Bacteriophage 186 complete genome."  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Xue Q., Egan J.B.;  
 RL EMBL: U32222; AAC34163.1; -;  
 SQ SEQUENCE 176 AA; 19643 MW; E53F5764AB650182 CRC64;

Query Match 47.3%; Score 43; DB 9; Length 176;  
 Best Local Similarity 75.0%; Pred. No. 33;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 VPTANAF 12  
 :|| |||||  
 Db 46 LPWLAWAF 53

RESULT 12  
 Q9LSPI PRELIMINARY; PRT; 344 AA.  
 ID Q9LSPI  
 AC Q9LSPI;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE PECTINESTERASE-LIKE PROTEIN.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;

RESULT 13  
ID 092MA2 PRELIMINARY; PRT: 392 AA.  
AC 092MA2:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE PUTATIVE PHOSPHOESTERASE AMINOTRANSFERASE PROTEIN (EC 2.6.1.52).  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21368234; PubMed=11474104;  
RA Gilbert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,  
Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Bothe G.,  
Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,  
Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,  
Glox S., Godrie T., Goffeau A., Golding B., Guzy J., Gurjal M.,  
Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,  
Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaure V.,  
Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,  
Ramsperger U., Surzycki R., Thebault P., Vandebol M.,  
Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;  
RT "The composite genome of the legume symbiont Sinorhizobium meliloti";  
RL Science 293:668-672(2001).  
DR EMBL; AL591791; CAC4310.1; -;  
KW Transferrase; Amino transferase; Complete proteome.  
SQ SEQUENCE 392 AA; 42336 MW; 1BFD16E10E39F78A CRC64;

Query Match 47.3%; Score 43; DB 10; Length 344;  
Best Local Similarity 62.5%; Pred. No. 63;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 7 WTMAAFHG 14  
|||::||  
DB 280 WTNMSYHG 287

Query Match 47.3%; Score 43; DB 16; Length 392;  
Best Local Similarity 75.0%; Pred. No. 71;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 VPWTAMAF 12  
|||::||  
DB 373 MPWTAMAF 380

RESULT 14  
ID 09A0U1 PRELIMINARY; PRT: 433 AA.  
AC 09A0U1:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE HYPOTHETICAL PROTEIN SPY0621.  
GN SPY0621.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
Primeaux G., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
Qian Y., Jia H.G., Najaf F.Z., Ren O., Zhu H., Song L., White J.,  
Ryan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL; AE006517; AAK33592.1; -;  
DR InterPro; IPR002819; HD.  
DR InterPro; IPR003607; HDC.  
DR Pfam; PF01966; HD: 1.  
DR SMART; SM00471; HDC: 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 433 AA; 50500 MW; 0A1FC654A87E758 CRC64;

Query Match 47.3%; Score 43; DB 16; Length 433;  
Best Local Similarity 53.8%; Pred. No. 78;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 VSRVPTAMAFHG 14  
: |||::||  
DB 36 IKOVPTTATFTFG 48

RESULT 15  
ID 09HYE2 PRELIMINARY; PRT: 441 AA.  
AC 09HYE2:  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE HYPOTHETICAL PROTEIN PA3464.  
GN PA3464.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Steyer C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltzy R., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
Barody L.L., Coulter S.N., Folger K.R., Kass A., Lardig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004767; AAG06852.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 441 AA; 49000 MW; 4923F37C2A044DA2 CRC64;

Query Match 47.3%; Score 43; DB 16; Length 441;

Best Local Similarity 85.7%; Pred. No. 80;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 6 PWTAWAF 12  
 |||||  
 Db 144 PWTAWRF 150

Search completed: August 15, 2002, 11:47:16  
 Job time: 527 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:38:24 ; Search time 73.69 seconds  
(without alignments)  
22.610 Million cell updates/sec

Title: US-09-613-092a-6

Perfect score: 87

Sequence: 1 RSXQHDLRAYGFWRLL 15

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A.Geneseq\_032802:\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	15	20	AAV30352
2	47	54.0	22	22	ABG19731
3	47	54.0	22	22	ABG19785
4	47	54.0	1007	22	ABG03083
5	44	50.6	267	22	ABG26189
6	44	50.6	1114	21	AA808517
7	44	50.6	1262	22	AAW79463
8	44	50.6	1289	22	AAW79447
9	44	50.6	1891	22	ABG04331
10	43	49.4	21	21	AAV75223
11	43	49.4	21	21	AAV75225

12	42	48.3	652	22	ABB67974	Drosophila melanog
13	42	48.3	951	20	AAV34536	Porphyromonas ging
14	42	48.3	953	20	AAV34403	Porphyromonas ging
15	41	47.1	111	22	ABG10728	Novel human diagno
16	41	47.1	111	22	ABG25967	Novel human diagno
17	41	47.1	588	22	AAU27909	Human contig polyp
18	41	47.1	651	22	ABG08359	Novel human diagno
19	40	46.0	87	22	ABG08529	Novel human diagno
20	40	46.0	249	21	AAG06462	Arabidopsis thalia
21	40	46.0	266	21	AAG06461	Arabidopsis thalia
22	40	46.0	311	21	AAG06460	Arabidopsis thalia
23	40	46.0	334	20	AAV32076	Bacillus cereus in
24	40	46.0	640	22	ABG10625	Novel human diagno
25	40	46.0	920	22	AAE09955	Methylomonas 16a n
26	39	44.8	198	22	AAU69448	Human purified sec
27	39	44.8	229	17	AAW05051	Human thrombopoiet
28	39	44.8	263	22	ABG04508	Novel human diagno
29	39	44.8	263	22	ABG17554	Novel human diagno
30	39	44.8	263	22	AAU32821	Novel human secret
31	39	44.8	282	17	AAW05045	Human thrombopoiet
32	39	44.8	282	17	AAW05045	Human thrombopoiet
33	39	44.8	286	16	AAW08027	Human MCPF-3. Hom
34	39	44.8	286	16	AAW05044	Human thrombopoiet
35	39	44.8	286	17	AAW05044	Human thrombopoiet
36	39	44.8	286	17	AAW05044	Human thrombopoiet
37	39	44.8	286	17	AAW05044	Human thrombopoiet
38	39	44.8	286	17	AAW05044	Human thrombopoiet
39	39	44.8	286	17	AAW05044	Human thrombopoiet
40	39	44.8	286	17	AAW05044	Human thrombopoiet
41	39	44.8	286	17	AAW05044	Human thrombopoiet
42	39	44.8	286	17	AAW05044	Human thrombopoiet
43	39	44.8	286	17	AAW05044	Human thrombopoiet
44	39	44.8	286	17	AAW05044	Human thrombopoiet
45	39	44.8	286	17	AAW05044	Human thrombopoiet

#### ALIGNMENTS

RESULT 1	AAV30352	standard: Peptide: 15 AA.
XX	AAV30352;	
AC	AAV30352;	
XX	09-NOV-1999 (first entry)	
DT	09-NOV-1999 (first entry)	
XX	Epitope derived from pneumococcal surface adhesion A protein.	
DE	Epitope derived from pneumococcal surface adhesion A protein.	
XX	Pneumococcal surface adhesion A protein; PsaA; monoclonal antibody;	
KW	vaccine; Streptococcus pneumoniae infection.	
KW	vaccine; Streptococcus pneumoniae infection.	
XX	Streptococcus pneumoniae.	
OS	Streptococcus pneumoniae.	
XX	WO9945121-A1.	
PN	WO9945121-A1.	
XX	10-SEP-1999.	
PD	10-SEP-1999.	
XX	26-FEB-1999; 99WO-US04326.	
PF	26-FEB-1999; 99WO-US04326.	
XX	02-MAR-1998; 98US-0076565.	
PR	02-MAR-1998; 98US-0076565.	
XX	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
PI	Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;	
PI	Zeller JL;	
XX	WPI: 1999-540849/45.	
DR	New peptides corresponding to Streptococcus pneumoniae PsaA, used	
XX	for treating or preventing Streptococcus pneumoniae infection in a	
PT	subject	
PT	subject	
XX	subject	

PS Claim 6; Page 43; 58pp; English.

XX AAY03051-54 represent immunogenic peptides which are derived from  
CC a pneumococcal surface adhesin A protein (PsaA). The specification  
CC describes monoclonal antibodies which bind epitopes of the PsaA protein  
CC (e.g. present sequence). The peptides can be used in vaccines to prevent  
CC Streptococcus pneumoniae infections. The antibodies of the invention  
CC can also be used to detect S. pneumoniae in a sample or individual.

XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 87; DB 20; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.9e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RSYOHLRAYGFWR 15  
| | | | | | | | | | | | | | |  
Db 1 rsyghdlraygfwr 15

RESULT 2  
ABG19731  
ID ABG19731 standard; Protein: 554 AA.  
XX  
AC ABG19731;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #19722.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR N-PSDB; AAS83918.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
PS Claim 20; SEQ ID NO 50090; 103pp; English.

XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG030377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 554 AA;

Query Match 54.0%; Score 47; DB 22; Length 554;  
Best Local Similarity 53.8%; Pred. No. 12;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 YOHLRAYGFWR 15  
| | | | | | | | | | | | | | |  
Db 7 ykshlaygawai 19

RESULT 3  
ABG19785  
ID ABG19785 standard; Protein: 554 AA.  
XX  
AC ABG19785;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #19776.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR N-PSDB; AAS83972.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
PS Claim 20; SEQ ID NO 50144; 103pp; English.

XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity



CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from W  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
XX  
Sequence 554 AA:  
50

Query Match	54.08;	Score 47;	DB 22;	Length 554;
Best Local Similarity	53.88;	Pred. No. 12;		
Matches 7; Conservative	2;	Mismatches 4;	Indels 0;	Gaps 0;

```
QY      3 YQHDLRAYGFWRL 15
          | : | | | | | :
Db      7 ykhsi1haygawai 19
```

RESULT 4  
ABG03083  
ID ABG03083 standard; Protein; 1007 AA.

AC	ABG03083;
XX	
DT	13-FEB-2002 (first entry)

XX Human; chromosome mapping; gene mapping; gene therapy; forensic  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
KW

OS	Homo sapiens.
XX	
PN	W0200175067-A2.

PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.

PR 23-AUG-2000; 200000  
XX  
PA (HYSE-) HYSEQ INC.

PT New Isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity -

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC

Query Match	54.0%;	Score 47;	DB 22;	Length 1007;
Best Local Similarity	53.8%;	Pred. No. 22;		
Matches 7; Conservative		2; Mismatches	4; Indels	0; Gaps 0;

QY	3	YQHDLRAYGEWRL	15
		:          :	
Db	7	Ykhs1haygawai	19

RESULT	5
ABG26189	
ID	ABG26189 standard; Protein; 267 AA

AC	ABG26189;
XX	
DT	18-FEB-2002 (first entry)

XX Human; chromosome mapping; gene mapping; gene therapy; forensic food supplement; medical imaging; diagnostic; genetic disorder.

OS	Homo sapiens.
XX	
PN	W0200175067-A2.

PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.

PR	23-AUG-2000; 20000
XX	
PA	(HYSE-) HYSEQ INC.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutation  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30337 represent novel human

CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 267 AA;

Query Match 50.6%; Score 44; DB 22; Length 267;  
 Best Local Similarity 53.8%; Pred. No. 18;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 RSYQHDLRAYGFW 13  
 ||| ||| |||  
 Db 120 rayradlrlegyw 132

# RESULT 6

ID AAB08517 standard; Protein; 1114 AA.

AC AAB08517;

DF 20-DEC-2000 (first entry)

DE Amino acid sequence of a beta-tubulin antigen.

KM Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;

OS chronic ear disease; autoimmune disease.

PN Homo sapiens.

PD WO200050593-A1.

PF 25-FEB-2000; 2000WO-US04795.

PR 25-FEB-1999; 9905-0121549.

PA (UYTE-) UNIV TENNESSEE RES CORP.

PI YOO TU;

DR WPI; 2000-558400/51.

PT New beta-tubulin antigen in the membranous structure of the inner ear,

PT reactive with antibodies of patients with Meniere's disease, for

PT diagnosing Meniere's disease and distinguishing this disease from other

PT autoimmune ear diseases -

PS Claim 3; Page 104-106; 115pp; English.

XX The present sequence represents a beta-tubulin antigen. The protein is

CC an antigen of the membranous structure of the inner ear protein, and

CC is reactive with antibodies from patients having Meniere's disease.

CC Meniere's disease is a chronic ear disease with unknown etiology.

CC Serum from patients suffering from this disease contain autoantibodies

CC against a 30 kDa cochlear protein antigen. The disease is believed to be

CC an autoimmune disease. The beta-tubulin antigen is useful as a target

CC substance in diagnosing or detecting Meniere's disease and in

CC distinguishing this disease from other autoimmune ear diseases.

SO Sequence 1114 AA;

Query Match 50.6%; Score 44; DB 21; Length 1114;  
 Best Local Similarity 58.3%; Pred. No. 76;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 YQHDLRAYGFW 14  
 ||| ||| |||  
 Db 32 yqhdrylygqwk 43

# RESULT 7

ID AAM78463 standard; Protein; 1262 AA.

AC AAM78463;

DF 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 1125.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KM tissue growth factor; immunomodulatory; cancer; leukaemia;

OS nervous system disorder; arthritis; inflammation.

PN Homo sapiens.

PD WO200157190-A2.

PF 09-AUG-2001.

PR 05-FEB-2001; 2001WO-US04098.

PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wehrman T, Goodrich R;

DR WPI; 2001-476283/51.

DR N-PSDB; AAK51596.

PT Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

PS Claim 20; Page 3356-3359; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666

CC (AAM80020) are omitted as the relevant pages from the sequence listing

CC were missing at the time of publication.

SO Sequence 1262 AA;

Query Match 50.6%; Score 44; DB 22; Length 1262;  
 Best Local Similarity 58.3%; Pred. No. 87;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 YQHDLRAYGFW 14  
 ||| ||| |||  
 Db 180 yqhdrylygqwk 191

RESULT 8  
AAM79447  
ID AAM79447 standard; Protein; 1289 AA.  
XX  
XX AAM79447:  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 3093.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
OS Homo sapiens.  
PN WO200157190-A2.  
PD  
XX 09-AUG-2001.  
PD  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI: 2001-476283/51.  
DR N-PSDB; AAK52580.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 250-251; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX  
SQ Sequence 1289 AA;

Query Match 50.6%; Score 44; DB 22; Length 1289;  
Best Local Similarity 58.3%; Pred. No. 89;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 YOHDLRAYGFW 14  
||| | | | | |  
Db 205 yqhyrylygwk 216

RESULT 9  
ABG04331  
ID ABG04331 standard; Protein; 1891 AA.  
XX  
XX ABG04331;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #4322.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
KW  
XX  
OS Homo sapiens.  
PN WO200175067-A2.  
PD  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI: 2001-639362/73.  
DR N-PSDB; AAS68518.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 20; SEQ ID NO 34690; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX  
SQ Sequence 1891 AA;

Query Match 50.6%; Score 44; DB 22; Length 1891;  
Best Local Similarity 53.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RSYOHLRAYGFW 13  
||:| ||| | :|  
Db 120 rayradrliegw 132

```
RESULT 10
AAV75223
ID AAV75223 standard; Protein; 459 AA.
XX
XX AAV75223;
AC
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 607 protein sequence SEQ ID NO:1920.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO957280-A2.
PD
PD 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR N-PSDB; AA253985.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX Claim 2; Page 967; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 459 AA:
SQ
```

```
Query Match 49.4%; Score 43; DB 21; Length 459;
Best Local Similarity 50.0%; Pred. No. 45;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 SYQHDLRAYGF 13
: | : | | | |
Db 416 ayrfdmgiygf 427
```

```
RESULT 11
AAV75225
ID AAV75225 standard; Protein; 459 AA.
XX
XX AAV75225;
AC
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 607 protein sequence SEQ ID NO:1924.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO957280-A2.
PD
PD 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR N-PSDB; AA253987.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX Claim 2; Page 969; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 459 AA:
SQ
```

```
Query Match 49.4%; Score 43; DB 21; Length 459;
Best Local Similarity 50.0%; Pred. No. 45;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 SYQHDLRAYGF 13
: | : | | | |
Db 416 ayrfdmgiygf 427
```

RESULT 12

ABB67974  
ID ABB67974 standard; Protein; 652 AA.  
XX  
XX ABB67974;  
AC  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 30714.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
PN WO200171042-A2.  
PD  
XX  
XX 27-SEP-2001.  
PE  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL12077.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 30714; 21pp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 652 AA;  
XQ

Query Match	48.3%	Score 42;	DB 22;	Length 652;
Best Local Similarity	53.8%	Pred. NO. 95;		
Matches	7;	Conservative	2;	Mismatches 4;
				Indels 0;
				Gaps 0.
QY	1 RSYQHDLRAYGFW 13			
	: :			
Db	557 rsfgyslctdyhfw 569			
RESULT	13			
AAV34536				
ID	AAV34536 standard; Protein; 951 AA.			
XX				
XX	AAV34536;			
XX				
XX	25-AUG-1999 (first entry)			
XX				
XX	Porphyromonas gingivalis protein PG67.			
DE				
XX				
XX	Porphyromonas gingivalis; PG; periodontal disease; gingivitis;			
KW	vaccline; antigenic.			
XX				
XX				

OS	Porphyromonas gingivalis.
XX	
PN	W09929870-A1.
XX	
PD	17-JUN-1999.
XX	
PF	10-DEC-1998; 98WO-A001023.
XX	
PR	04-AUG-1998; 98AU-0005028.
PR	10-DEC-1997; 97AU-0000839.
PR	31-DEC-1987; 97AU-0001182.
PR	30-JAN-1998; 98AU-0001546.
PR	10-MAR-1998; 98AU-0002264.
PR	09-APR-1998; 98AU-0002911.
PR	23-APR-1998; 98AU-0003128.
PR	05-MAY-1998; 98AU-0003338.
PR	22-MAY-1998; 98AU-0003654.
PR	29-JUL-1998; 98AU-0004917.
XX	
PPA	(CSLC-) CSL LTD.
XX	
PI	Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI	Ross BC, Rothel LJ, Webb EA;
XX	
DR	WPI. 1999-385613/32.
DR	N-PSDB; AAX91754.

```

PT Antigenic Porphyromonas gingivalis peptides for preventing
PI gingivitis
XX
XX Claim 1; Page 526-527; 588bp; English.
PS
XX
XX
CC AAY91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
CC
XX
XX Sequence 951 AA;
SQ
Query Match 48.3%; Score 42; DB 20; Length 951;
Best Local Similarity 56.2%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 5; Indels 2; Gaps 1;
OY 2 SYOHDL--RAYGFWRLL 15
||| | | |||| |
DB 599 sympdfktrrygfwel 614
RESULT 14
AAY34403
ID AAY34403 standard; Protein: 953 AA.
XX
XX AAY34403;
AC
XX
XX 25-AUG-1999 (first entry)
DT
XX
XX Porphyromonas gingivalis protein PG67.
DE
XX
XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic.
XX
XX Porphyromonas gingivalis.
OS
XX
XX WO9929870-A1.
PN
XX
XX 17-JUN-1999.
PD
XX

```

PF 10-DEC-1998; 98MO-AU01023.  
XX  
XX 04-AUG-1998; 98AU-0005028.  
PR 10-DEC-1997; 97AU-0000839.  
PR 31-DEC-1997; 97AU-0001182.  
PR 30-JAN-1998; 98AU-0001546.  
PR 10-MAR-1998; 98AU-0002264.  
PR 09-APR-1998; 98AU-0002911.  
PR 23-APR-1998; 98AU-0003128.  
PR 05-MAY-1998; 98AU-0003338.  
PR 22-MAY-1998; 98AU-0003654.  
PR 29-JUL-1998; 98AU-0004917.  
XX  
XX (CSLC-) CSL LTD.  
XX  
PI Ajius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
PI Ross BC, Rothel LJ, Webb EA;  
XX  
XX WPI: 1999-385613/32.  
DR N-PSDB; AAX91621.  
XX  
XX Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis  
XX  
PS Claim 1; Page 373-375; 588pp; English.  
XX  
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the  
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
CC activity with a vaccine mechanism of action. The PG polypeptides can be  
CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
CC be used to detect Porphyromonas gingivalis in standard hybridisation  
CC assays. Porphyromonas gingivalis is involved in periodontal disease  
CC especially gingivitis.  
XX  
XX Sequence 953 AA;  
SQ  
Query Match 48.3%; Score 42; DB 20; Length 953;  
Best Local Similarity 56.2%; Pred. No. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 5; Indels 2; Gaps 1;  
QY 2 SYQHDL--RAYGFWRL 15  
Db 601 sympdfkrryglw 616  
RESULT 15  
ABG10728  
ID ABG10728 standard; Protein: 111 AA.  
XX  
XX ABG10728;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX Novel human diagnostic protein #10719.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001MO-US08631.  
PE  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA

XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS74915.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 20; SEQ ID NO 41087; 103pp; English.  
PS  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 111 AA;  
SQ  
Query Match 47.1%; Score 41; DB 22; Length 111;  
Best Local Similarity 38.5%; Pred. No. 23;  
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 1 RSYQHDLRAYGFW 13  
Db 51 rgrndirgfaqw 63

Search completed: August 15, 2002, 11:38:26  
Job time: 267 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:38:58 ; Search time 26.88 Seconds  
(without alignments)  
13.630 Million cell updates/sec

Title: US-09-613-092a-6  
Perfect score: 87  
Sequence: 1 RSYQHDLRAYGFWRL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents -AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCFUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfillseq1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	44.8	265	1	US-08-413-803-27
2	39	44.8	265	5	PCT-US95-03776-29
3	39	44.8	286	1	US-08-321-488A-27
4	39	44.8	2104	2	US-08-808-793-4
5	39	44.8	2104	3	US-08-772-512A-4
6	39	44.8	2105	2	US-08-808-793-3
7	39	44.8	2105	3	US-08-772-512A-3
8	38	43.7	66	4	US-08-936-165A-376
9	38	43.7	148	3	US-08-946-329A-65
10	38	43.7	320	1	US-08-362-739-2
11	38	43.7	1111	1	US-08-317-450B-15
12	38	43.7	1111	4	US-08-800-593-15
13	38	43.7	1171	1	US-08-445-135-1
14	38	43.7	1193	4	US-08-817-450B-13
15	38	43.7	1193	4	US-08-800-593-13
16	37	42.5	76	4	US-09-227-357-456
17	37	42.5	187	4	US-09-247-155-102
18	37	42.5	567	2	US-08-504-459-6
19	37	42.5	567	2	US-08-504-459-6
20	37	42.5	616	1	US-08-149-100-2
21	37	42.5	695	6	5460961-5
22	37	42.5	877	1	US-08-430-925A-4
23	37	42.5	940	2	US-08-938-365-4
24	37	42.5	941	1	US-08-343-760A-2
25	36	41.4	74	2	US-08-743-130A-41
26	35.5	40.8	67	2	US-08-248-839C-100
27	35.5	40.8	607	3	US-08-781-891-75

28	35	40.2	53	4	US-09-187-789-65	Sequence 65, Appl
29	35	40.2	140	1	US-07-946-421-24	Sequence 24, Appl
30	35	40.2	140	1	US-07-946-421-28	Sequence 28, Appl
31	35	40.2	201	2	US-08-716-317-1	Sequence 1, Appl
32	35	40.2	214	4	US-08-861-774E-74	Sequence 74, Appl
33	35	40.2	214	4	US-08-861-774E-78	Sequence 78, Appl
34	35	40.2	242	2	US-08-553-497A-26	Sequence 26, Appl
35	35	40.2	242	2	US-08-553-497A-28	Sequence 28, Appl
36	35	40.2	242	4	US-09-019-095A-23	Sequence 23, Appl
37	35	40.2	244	2	US-08-553-497A-20	Sequence 20, Appl
38	35	40.2	244	2	US-08-553-497A-22	Sequence 22, Appl
39	35	40.2	246	2	US-08-553-497A-24	Sequence 24, Appl
40	35	40.2	260	4	US-09-187-789-2	Sequence 2, Appl
41	35	40.2	463	4	US-09-163-444-2	Sequence 2, Appl
42	35	40.2	474	3	US-09-022-669-2	Sequence 2, Appl
43	35	40.2	485	2	US-08-446-803-1	Sequence 1, Appl
44	35	40.2	485	2	US-08-861-837-1	Sequence 1, Appl
45	35	40.2	485	2	US-08-600-908A-12	Sequence 12, Appl

## ALIGNMENTS

RESULT 1  
US-08-413-803-27  
Sequence 27, Application US/08413803  
Patent No. 5766581  
GENERAL INFORMATION:  
APPLICANT: Bartley, Timothy D.  
APPLICANT: Bogenberger, Jakob M.  
APPLICANT: Bosselman, Robert A.  
APPLICANT: Hunt, Pamela  
APPLICANT: Kinsler, Olaf B.  
APPLICANT: Samal, Babu B.  
TITLE OF INVENTION: METHODS FOR TREATING MAMMALS WITH  
TITLE OF INVENTION: MONO-PEGYLATED PROTEINS THAT STIMULATE MEGAKARYOCYTE  
GROWTH AND DIFFERENTIATION  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: 1840 DeHavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: US  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/413, 803  
APPLICATION NUMBER: US/08/413, 803  
FILING DATE: 30-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/221,768  
FILING DATE: 31-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/252,628  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/321,488  
FILING DATE: 12-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/347,780  
FILING DATE: 30-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Cook Ph.D., Robert R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-290D  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-413-803-27

Query Match 44.8%; Score 39; DB 1; Length 265;  
Best Local Similarity 58.3%; Pred. No. 24;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 SYQHDLRAGFW 13  
|:|:|:|:|:|:  
Db 130 SFQHLRGKDFW 141

RESULT 2  
PCT-US95-03776-29  
Sequence 29, Application PC/TUS9503776  
GENERAL INFORMATION:  
APPLICANT: AMGEN INC.  
TITLE OF INVENTION: Compositions and Methods for Stimulating  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Denavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/03776  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cook, Robert R.  
REFERENCE/DOCKET NUMBER: A-290-C  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-03776-29

Query Match 44.8%; Score 39; DB 5; Length 265;  
Best Local Similarity 58.3%; Pred. No. 24;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 SYQHDLRAGFW 13  
|:|:|:|:|:|:  
Db 130 SFQHLRGKDFW 141

RESULT 3  
US-08-321-488A-27  
Sequence 27, Application US/08321488A  
GENERAL INFORMATION:  
APPLICANT: Bartley, Timothy D.  
APPLICANT: Bogenberger, Jakob M.  
APPLICANT: Bosselman, Robert A.  
APPLICANT: Hunt, Pamela  
APPLICANT: Kinstler, Olaf B.  
APPLICANT: Samal, Badru B.  
TITLE OF INVENTION: MONO-PEGYLATED PROTEINS THAT STIMULATE  
TITLE OF INVENTION: MEGAKARYOCYTE GROWTH AND DIFFERENTIATION

NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: 1840 Denavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/321,488A  
FILING DATE: 12-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/252,628  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/221,768  
FILING DATE: 31-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Cook, Robert R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-290B  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 266 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-321-488A-27

Query Match 44.8%; Score 39; DB 1; Length 286;  
Best Local Similarity 58.3%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 SYQHDLRAGFW 13  
|:|:|:|:|:|:  
Db 151 SFQHLRGKDFW 162

RESULT 4  
US-08-808-793-4  
Sequence 4, Application US/08808793  
Patent No. 5858713  
GENERAL INFORMATION:  
APPLICANT: Soderlund, David M.  
APPLICANT: Ingles, Patricia J.  
TITLE OF INVENTION: CALCIUM PERMEABLE INSECT SODIUM CHANNELS  
TITLE OF INVENTION: AND USE THEREOF  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: USA  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/808,793  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/034,361



Db 383 YOHVLQAGPWHM 395

## RESULT 7

US-08-772-512A-3  
; Sequence 3, Application US/08772512A

; Patent No. 6022705

; GENERAL INFORMATION:

; APPLICANT: Soderlund, David M.

; APPLICANT: Knippler, Douglas C.

; APPLICANT: Ingles, Patricia J.

; TITLE OF INVENTION: INSECT SODIUM CHANNELS FROM

; TITLE OF INVENTION: INSECTICIDE-SUSCEPTIBLE AND INSECTICIDE-RESISTANT HOUSE

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NIXON, HARGRAVE, DEVANS &amp; DOYLE LLP

; STREET: P.O. Box 1051, Clinton Square

; CITY: Rochester

; STATE: New York

; COUNTRY: USA

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/772,512A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/608,618

; FILING DATE: 01-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Braman, Susan J.

; REGISTRATION NUMBER: 34,103

; REFERENCE/DOCKET NUMBER: 19603/601(CRPD-1657)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 716-263-1636

; TELEFAX: 716-263-1600

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2105 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-772-512A-3

Query Match 44.8%; Score 39; DB 3; Length 2105;

Best Local Similarity 53.8%; Pred. No. 2.2e+02;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 YOHDLRAYGFWR 15

Db 383 YOHVLQAGPWHM 395

## RESULT 8

US-08-936-165A-376

; Sequence 376, Application US/08936165A

; Patent No. 6348582

; GENERAL INFORMATION:

; APPLICANT: Black, Michael

; APPLICANT: Burnham, Martin

; APPLICANT: Hodgson, John

; APPLICANT: Knowles, David

; APPLICANT: Lonetto, Michael

; APPLICANT: Nicholas, Richard

; APPLICANT: Pratt, Julie

; APPLICANT: Reichard, Richard

; APPLICANT: Rosenberg, Martin

; APPLICANT: Ward, Judith

; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,

; TITLE OF INVENTION: Polypeptides and Their Uses

; NUMBER OF SEQUENCES: 534

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmtlKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY: USA

; ZIP: 19406-0939

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/936,165A

; FILING DATE: 24-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/027,032

; FILING DATE: 24-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Gimmil, Edward R

; REGISTRATION NUMBER: 38,891

; REFERENCE/DOCKET NUMBER: P50549

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-270-4478

; TELEFAX: 610-270-5090

; TELEX:

; INFORMATION FOR SEQ ID NO: 376:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 66 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-936-165A-376

Query Match 43.7%; Score 38; DB 4; Length 66;

Best Local Similarity 50.0%; Pred. No. 8.2;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 SYOHLRAYGFWR 15

Db 25 SYEPIENAGTWR 38

## RESULT 9

US-08-946-329A-65

; Sequence 65, Application US/08946329A

; Patent No. 6057091

; GENERAL INFORMATION:

; APPLICANT: Beachy, Philip A.

; APPLICANT: Porter, Jeffrey A.

; TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES

; NUMBER OF SEQUENCES: 109

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: USA

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: Windows 95

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

```

: APPLICATION NUMBER: US/08/946,329A
: FILING DATE: 07-OCT-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/061,323
: FILING DATE: 07-OCT-1996
: APPLICATION NUMBER: 08/729,743
: FILING DATE: 10-JUL-1996
: APPLICATION NUMBER: 08/567,357
: FILING DATE: 04-DEC-1995
: APPLICATION NUMBER: 08/349,498
: FILING DATE: 02-DEC-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 07265/140001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619/678-5070
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 65:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 148 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-946-329A-65

```

```

Query Match      43.7%; Score 38; DB 3; Length 148;
Best Local Similarity 53.8%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY      2 SYQHDLRVGFWR 14
      1 111 1 11
DB      105 SODHDLGRGKWR 117

```

```

RESULT 10
US-08-362-739-2
: Sequence 2, Application US/08362739
: Patent No. 5658757
: GENERAL INFORMATION:
: APPLICANT: Haake, David A.
: APPLICANT: Bianco, David R.
: APPLICANT: Champion, Cheryl I.
: APPLICANT: Lovett, Michael A.
: APPLICANT: Miller, James N.
: TITLE OF INVENTION: CLONED Leptospira OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Spensley Horn Jubas & Lubitz
: STREET: 1880 Century Park East, Suite 500
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90067
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/362,739
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/040,747
: FILING DATE: 31-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Wetherell, Jr. Ph.D., John R.,
: REGISTRATION NUMBER: 31,678
: REFERENCE/DOCKET NUMBER: PD-2097
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 455-5100

```

```

: TELEFAX: (619) 455-5110
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 320 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-362-739-2

```

```

Query Match      43.7%; Score 38; DB 1; Length 320;
Best Local Similarity 46.2%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      3 YQHDLRVGFWR 15
      1 1 1111
DB      131 YESDFGKFFWRV 143

```

```

RESULT 11
US-08-317-450B-15
: Sequence 15, Application US/08317450B
: Patent No. 5660982
: GENERAL INFORMATION:
: APPLICANT: Tryggvason, Karl
: APPLICANT: Kallunki, Pekka
: APPLICANT: Pyke, Charles
: TITLE OF INVENTION: Laminin Chains: Diagnostic and
: TITLE OF INVENTION: Therapeutic Use
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BANNER & ALLEGRETTI, LTD.
: STREET: Ten South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/317,450B
: FILING DATE: 04-OCT-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Chao, Mark
: REGISTRATION NUMBER: 37,293
: REFERENCE/DOCKET NUMBER: 94,778
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-715-1000
: TELEFAX: 312-715-1234
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1111 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-317-450B-15

```

```

Query Match      43.7%; Score 38; DB 1; Length 1111;
Best Local Similarity 87.5%; Pred. No. 17e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 RSYQHDR 8
      11111 11
DB      843 RSYQHSIR 850

```

```

RESULT 12
US-08-800-593-15

```

```
; Sequence 15, Application US/08800593
; Patent No. 6143505
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-800-593-15

Query Match          43.7%; Score 38; DB 4; Length 1111;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RSYQHDLR 8
    |||||
Db 843 RSYQHSR 850

RESULT 13
; Sequence 1, Application US/08445135
; Patent No. 5658789
; GENERAL INFORMATION:
; APPLICANT: Quaranta, Vito
; APPLICANT: Hornia, Marketta
; TITLE OF INVENTION: Promotion of Epithelial Cell Adhesion
; TITLE OF INVENTION: and Hemidesmosome Assembly by a Laminin-Like Molecule
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,135
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/151,134
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: DESMOS.002CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; US-08-445-135-1

Query Match          43.7%; Score 38; DB 1; Length 1171;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RSYQHDLR 8
    |||||
Db 822 RSYQHSR 829

RESULT 14
; Sequence 13, Application US/08317450B
; Patent No. 5660982
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: Ten South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,450B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
```

; INFORMATION FOR SEQ ID NO: 13:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1193 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-317-450B-13

Query Match 43.7%; Score 38; DB 1; Length 1193;  
 Best Local Similarity 87.5%; Pred. NO. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RSYQHDLR 8  
 |||||  
 DB 843 RSYQHSLR 850

RESULT 15  
 US-08-800-593-13  
 ; Sequence 13, Application US/08800593  
 ; Patent No. 6143505  
 ; GENERAL INFORMATION:  
 ; APPLICANT: trygvason, Karl  
 ; APPLICANT: Kallunki, Pekka  
 ; APPLICANT: Pyke, Charles  
 ; TITLE OF INVENTION: Laminin Chains: Diagnostic and  
 ; TITLE OF INVENTION: Therapeutic Use  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
 ; STREET: 300 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/800,593  
 ; FILING DATE: 18-FEB-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/317,450  
 ; FILING DATE: 04-OCT-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Chao, Mark  
 ; REGISTRATION NUMBER: 37,293  
 ; REFERENCE/DOCKET NUMBER: 94,778-B  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-913-0001  
 ; TELEFAX: 312-913-0002  
 ; INFORMATION FOR SEQ ID NO: 13:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1193 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-800-593-13

Query Match 43.7%; Score 38; DB 4; Length 1193;  
 Best Local Similarity 87.5%; Pred. NO. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RSYQHDLR 8  
 |||||  
 DB 843 RSYQHSLR 850

Search completed: August 15, 2002, 11:38:58  
 Job time: 264 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:39:40 ; Search time 35.58 Seconds  
(without alignments)  
40.510 Million cell updates/sec

Title: US-09-613-092a-6  
Perfect score: 87  
Sequence: 1 RSYQHDLRAYGFWRLL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR-71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	49.4	118	2	G81992
2	43	49.4	118	2	G81221
3	43	49.4	188	2	T02822
4	43	49.4	459	2	DB1950
5	42	48.3	444	2	T24844
6	42	48.3	887	2	T03939
7	42	48.3	994	1	SYNCLM
8	41	47.1	188	2	D82248
9	41	47.1	817	2	T43659
10	40	46.0	223	2	A84436
11	40	46.0	274	2	F96961
12	40	46.0	330	2	B96027
13	40	46.0	371	2	T21153
14	40	46.0	414	2	T19912
15	40	46.0	447	2	E72320
16	40	46.0	451	2	T15718
17	40	46.0	516	2	D96682
18	40	46.0	912	2	T31223
19	40	46.0	1177	2	AC0284
20	39.5	45.4	309	2	A95988
21	39.5	45.4	471	2	D95940
22	39	44.8	81	2	A75372
23	39	44.8	103	2	A53461
24	39	44.8	206	2	S40829
25	39	44.8	206	2	H91229
26	39	44.8	206	2	G86076
27	39	44.8	286	2	A55530
28	39	44.8	313	2	AF1973
29	39	44.8	330	2	B97368

30	39	44.8	330	2	AB2586	pfkB family carboh
31	39	44.8	335	2	I36790	hypothetical prote
32	39	44.8	340	2	C83827	polysaccharin bios
33	39	44.8	433	1	B65058	f1c protein homol
34	39	44.8	502	2	B75287	sensor histidine k
35	39	44.8	505	2	F71266	probable tpr prote
36	39	44.8	514	2	H87465	sodium-galactoside
37	39	44.8	700	2	S09699	blb protein - fruI
38	39	44.8	736	2	T06757	hypothetical prote
39	39	44.8	978	2	H81311	transcription-repa
40	39	44.8	1055	2	T31111	AtPase I (EC 3.6.1
41	39	44.8	1135	1	JQ1928	G2-G1 polypeptid
42	39	44.8	2108	2	S72458	sodium channel pro
43	38.5	44.3	159	2	AH2118	bacterioferritin c
44	38.5	44.3	307	2	F97639	hypothetical prote
45	38.5	44.3	307	2	AG2862	conserved hypothet

## ALIGNMENTS

```

RESULT 1
G81992
NDH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain A NMA0019 [similarity] - Neisser
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: G81992
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
Nure 404, 502-506, 2000
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: AB1775; MUID:20222556
A:Accession: G81992
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-118 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83339.1; PID:g737
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: nuoa; NMA0019
C:Superfamily: NDH dehydrogenase (ubiquinone) chain 3
C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 49.4%; Score 43; DB 2; Length 118;
Best Local Similarity 70.0%; Pred. No. 4.5;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DLRAYGFWRLL 15
DB 86 DLGAYGFWRSM 95

RESULT 2
G81221
NDH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain A NMB0241 [similarity] - Neisser
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81221
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
R:Teitelin, H.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: G81221
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-118 <TEP>
A:Cross-references: GB:AE002381; GB:AE002098; NID:g7225455; PIDN:AAF40695.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:

```

```

A:Gene: NMB0241;
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 3
C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match          49.4%; Score 43; DB 2; Length 118;
Best Local Similarity 70.0%; Pred. No. 4.5;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DLRAYGFWRL 15
    |||||
    |||||
Db 86 DLGAGFWSM 95

RESULT 3
T02812
Probable membrane protein L2759.14 [Imported] - Leishmania major (strain Friedlin)
C:Species: Leishmania major
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 19-May-2000
C:Accession: B61459; T02822
R:Myler, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999
A:Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c
A:Reference number: A81455; MUID:99178987
A:Accession: B81459
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <P>
A:Cross-references: GB:AE001274; NID:93264850; PTDN:AAC24644.1; PID:g1407729; GSPDB:GN00
C:Experimental source: strain MHOM/IL/81/Friedlin
C:Genetics:
A:Gene: L2759.14
A:Map position: 1
C:Keywords: transmembrane protein

Query Match          49.4%; Score 43; DB 2; Length 168;
Best Local Similarity 77.8%; Pred. No. 7.3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 HDLRAYGFW 13
    |||
    |||
Db 46 HDLMAIGFW 54

RESULT 4
D81950
Probable transmembrane efflux protein NMA1022 [Imported] - Neisseria meningitidis (stra
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D81950
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:2022556
A:Accession: D81950
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-459 <P>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PTDN:CAB84291.1; PID:g73797729
C:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1022
C:Superfamily: conserved hypothetical protein H11612

Query Match          49.4%; Score 43; DB 2; Length 459;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 SYOHDLRAYGFW 13
    :|:|:|

```

```

Db          416 AYRDMGIFYGM 427

RESULT      5
T24844
hypothetical protein T11F9.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24844
R:Lennard, N.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19941
A:Accession: T24844
A>Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-444 <TLL>
A:Cross-references: EMBL:Z74042; PIDN:CAA98536.1; GSPDB:GN00023; CESP:T11F9.1
A:Experimental source: clone T11F9
A:Genetics:
A:Gene: CESP:T11F9.1
A:Map position: 5
A:Introns: 19/1; 61/2; 130/3; 177/3; 347/3; 375/1; 413/3

Query Match          48.3%; Score 42; DB 2; Length 444;
Best Local Similarity 42.9%; Pred. No. 27;
Matches 9; Conservative 3; Mismatches 3; Indels 6; Gaps 1;

OY      1 RSYOHLRA-----YGFWRL 15
        | | | | |
Db      189 RYQLNLSALRYELTYGM 209

RESULT      6
T03939
potassium channel protein - maize
C:Species: Zea mays (maize)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-Jul-2000
C:Accession: T03939
R:Hotl, S.; Dreyer, I.; Dietrich, P.; Becker, D.; Mueller-Roeber, B.; Hedrich, R.
Proc. Natl. Acad. Sci. U.S.A. 94, 4806-4810, 1997
A>Title: Molecular basis of plant-specific acid activation of K+ uptake channels.
A:Reference number: Z14922; MUID:97272307
A:Accession: T03939
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-887 <HOT>
A:Cross-references: EMBL:Y07632; NID:g2104907; PIDN:CAA68912.1; PID:g2104908
A:Experimental source: Strain U51311B, sub-species Apache, coleoptile
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
C:Keywords: potassium channel; transmembrane protein; transport protein; voltage-gate
F:561-593/Domain: ankyrin repeat homology <AN1>
F:658-690/Domain: ankyrin repeat homology <AN2>

Query Match          48.3%; Score 42; DB 2; Length 887;
Best Local Similarity 56.2%; Pred. No. 55;
Matches 9; Conservative 1; Mismatches 0; Indels 6; Gaps 1;

OY      6 DLRAYGF-----WRL 15
        |||:| | |
Db      163 DLRSYGFNMLRL 178

RESULT      7
SYNCLM
leucine--tRNA ligase (EC 6.1.1.4) precursor, mitochondrial - Neurospora crassa
N:Alternate names: Leucyl-tRNA synthetase
C:Species: Neurospora crassa
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A33474
R:Chow, C.M.; Metzenberg, R.L.; Rajbhandary, U.L.
Mol. Cell. Biol. 9, 4631-4644, 1989

```

```

A:Title: Nuclear gene for mitochondrial leucyl-tRNA synthetase of Neurospora crassa: iso
A:Reference number: A33474; MUID:90097874
A:Accession: A33474
A:Molecule type: DNA
A:Residues: 1-994 <CHD>
A:Cross-references: GB:M30472; NID:g168829; PIDN:AAA3359.1; PID:g168830
C:Genetics:
A:Gene: leu-5
C:Superfamily: leucine--tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; mitochondrion; protein biosynthesis;
F:1-52/Domin: transit peptide (mitochondrion) #status predicted <TNP>
F:53-994/Product: leucine-tRNA ligase #status predicted <MAT>
F:100-103/Region: ATP binding #status predicted
F:503-550/Region: zinc finger CCCC motif

OY      5 HDLRAYGFWR 14
      ||| : ||:
Db      403 HDLRHAFWK 412

RESULT      8
D82248
RNA polymerase sigma-70 factor, ECF subfamily VC1045 [imported] - Vibrio cholerae (strain
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82248
R:Heidelberg, U.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: D82248
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-188 <HEI>
A:Cross-references: GB:AE004186; GB:AE003852; NID:g9655507; PIDN:AAF94204.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1045
A:Map position: 1

Query Match      47.1%; Score 41; DB 2; Length 188;
Best Local Similarity 46.2%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY      1 RSYOHDLRAYGFW 13
      | : | | | : |
Db      32 RAYHDLRYAYW 44

RESULT      9
T43659
trehalose-phosphatase (EC 3.1.3.12) [similarity] - fission yeast (Schizosaccharomyces pombe
N:Alternate names: trehalose-6P phosphatase
C:Species: Schizosaccharomyces pombe
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43659; T38002; T43056
R:Canasdo, J.; Vicente, J.; Soto, T.; Franco, A.; Castillo, R.; Gacto, M.
submitted to the EMBL Data Library, June 1999
A:Description: Isolation and characterization of trehalose-6P phosphatase disrupted muta
A:Reference number: 222608
A:Accession: T43659
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-817 <CAN>
A:Cross-references: EMBL:AJ242743; PIDN:CAB45142.1

```

```

A:Experimental source: strain 972h(-)
R:Oliver, K.; Harris, D.; Wood, V.; Barrrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, June 1997
A:Reference number: 221760
A:Accession: T38002
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-817 <OL1>
A:Cross-references: EMBL:Z97209; PIDN:CAB10126.1; GSPDB:GN00066; SPDB:SPAC19612.15c
A:Experimental source: strain 972h-; cosmid cl9612
R:Yoshioke, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4: 363-369, 1997
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A:Reference number: Z17323; MUID:98162722
A:Accession: T43056
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 505-817 <YOS>
A:Cross-references: EMBL:DB8225; NID:91749657; PIDN:BAAL3886.1; PID:91749658
A:Experimental source: strain PR745
C:Genetics:
A:Gene: tppl; SPAC19612.15c
A:Map position: 1
C:Function:
A:Description: catalyzes the hydrolysis of trehalose 6-phosphate to trehalose
A:Pathway: trehalose biosynthesis
C:Keywords: phosphoric monoester hydrolase; trehalose biosynthesis

Query Match 47.1%; Score 41; DB 2; Length 817;
Best Local Similarity 57.1%; Pred. No. 75;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 RSYOHLRAYGFWR 14
||| ||| ||
Db 362 RSPQRFLETPYEW 375

RESULT 10
A84436
probable glutathione S-transferase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence.revision 02-Feb-2001 #text.change 16-Feb-2001
C:Accession: A84436
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: A84436
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-223 <STO>
A:Cross-references: GB:AE002093; NID:93894170; PIDN:AAC78520.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g02380
A:Map position: 2
C:Superfamily: plaiace glutathione transferase

Query Match 46.0%; Score 40; DB 2; Length 223;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 YOHDLRAYGFWR 14
||| ||| ||
Db 8 YQAKLKLYSWR 19

RESULT 11
F96961
bacitracin resistance protein (bach) [imported] - Clostridium acetobutylicum

```

C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: F96961  
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: F96961  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AKK78481.1; PID:g15023362; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0501  
C:Superfamily: Escherichia coli bacitracin resistance protein bacA

Query Match 46.0%; Score 40; DB 2; Length 274;  
Best Local Similarity 41.7%; Pred. No. 35;  
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 4 OHDLRAYGFWR 15  
|||:|:|:|:|:  
Db 250 RHDFKAFGWYRI 261

RESULT 12  
B96027  
probable aliphatic sulfonates uptake ABC transporter periplasmic solute-binding protein  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: B96027  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:2136508; PMID:11481431  
A:Accession: B96027  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-330 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CAC49882.1; PID:g15141370; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymb  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Punler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
Pelat, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Jelaure,  
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMD20570  
A:Genome: plasmid

Query Match 46.0%; Score 40; DB 2; Length 330;  
Best Local Similarity 50.0%; Pred. No. 43;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 RSYOHDLRAYGFWR 14  
|||:|:|:|:|:  
Db 179 RSNQVDIAAGVYQW 192

RESULT 13  
T21153  
hypothetical protein F20E11.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T21153

R:Ainscough, R.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19383  
A:Accession: T21153  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-371 <RTL>  
A:Cross-references: EMBL:Z61508; PIDN:CAB04139.1; GSPDB:GN00023; CESP:F20E11.6  
A:Experimental source: clone F20E11  
C:Genetics:  
A:Gene: CESP:F20E11.6  
A:Map position: 5  
A:Introns: 173/2, 211/3, 247/2

Query Match 46.0%; Score 40; DB 2; Length 371;  
Best Local Similarity 87.5%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 RAYGFWR 15  
|||||:  
Db 70 RAYGFWR 77

RESULT 14  
T19912  
hypothetical protein C43F9.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T19912  
R:Mortimore, B.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19195  
A:Accession: T19912  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-414 <WIL>  
A:Cross-references: EMBL:Z82262; PIDN:CAB05151.1; GSPDB:GN00022; CESP:C43F9.4  
A:Experimental source: clone C43F9  
C:Genetics:  
A:Gene: CESP:C43F9.4  
A:Map position: 4  
A:Introns: 53/3; 81/2; 111/3; 134/3; 164/3; 219/1; 346/1

Query Match 46.0%; Score 40; DB 2; Length 414;  
Best Local Similarity 60.0%; Pred. No. 54;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 4 OHDLRAYGFWR 13  
|||:|:|:|:|:  
Db 21 KHSLMFYGFW 30

RESULT 15  
E72320  
hypothetical protein TM0884 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: E72320  
R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316  
A:Accession: E72320  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-447 <ARN>  
A:Cross-references: GB:AE001754; GB:AE000512; NID:q4981417; PIDN:PAD35965.1; PID:q498  
A:Experimental source: strain MSB8  
C:Genetics:

A:Gene: TM0884

Query Match 46.0%; Score 40; DB 2; Length 447;  
 Best Local Similarity 66.7%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 5 HDLRAYCFW 13  
 |||||  
 Db 11 HDEAAYDFW 19

Search completed: August 15, 2002, 11:39:41  
 Job time: 202 sec

**THIS PAGE BLANK (USPTO)**

QY 1 RSYOHDLRAYGFW 14  
 ||:| || || ||  
 Db 362 RSYOHDLRAYGFW 375

RESULT 15

09FVG6 PRELIMINARY; PRT; 854 AA.  
 ID 09FVG6  
 AC 09FVG6;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TRANSPOSASE.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Poaceae; Andropogoneae; Zea.  
 ON NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W22: TRANSPOSON-DOPPIA4;  
 RA Bercury S.D., Walker E.L.;  
 RT "Molecular analysis of the structure and function of the DopPIA  
 RT transposable element of maize."  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF187822; AAC17043.1; -;  
 SQ SEQUENCE 854 AA; 99057 MW; C3F745C5EB706EA CRC64;

Query Match

Best Local Similarity 47.1%; Score 41; DB 10; Length 854;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 HDLRLAYGFW 13  
 || || || ||  
 Db 425 HDLRLAYGFW 433

Search completed: August 15, 2002, 11:47:17  
 Job time: 528 sec

QY 2 SYOHDLRAYGFV 13  
 :|:| | | |  
 DB 202 AVEHKERFYGV 213

RESULT 12  
 052212 PRELIMINARY: PRT: 404 AA.

ID 052212  
 AC 052212:  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TNPA.  
 GN Serratia marcescens.  
 OS Plasmid R471a.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=75096689; PubMed=1089756;  
 RA Hedges R.W., Rodriguez-Lemoin V., Datta N.;  
 RT "R factors from Serratia marcescens.";  
 RL J. Gen. Microbiol. 86:88-92(1975).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93374835; PubMed=8366028;  
 RA Ho C., Kulaeva O.I., Levine A.S., Woodgate R.;  
 RT "A rapid method for cloning mutagenic DNA repair genes: isolation of  
 RT unu-complementing genes from multidrug resistance plasmids R391,  
 RT R446b, and R471a.";  
 RL J. Bacteriol. 175:5411-5419(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98202731.  
 RA Kulaeva O.I., Koonin E.V., Wootton J.C., Levine A.S., Woodgate R.;  
 RT "Unusual insertion element polymorphisms in the promoter and  
 RT terminal regions of the mucAB-like genes of R471a and R446b.";  
 RL Mutat. Res. 397:247-262(1998).  
 DR EMBL: AF027768; AAC82523.1; -;  
 DR InterPro: IPR002560; Transposase\_12.  
 DR Pfam: PF01610; Transposase\_12; 1.  
 KM Plasmid.  
 SQ SEQUENCE 404 AA: 47498 MW; 2B130E43026E5404 CRC64;

Query Match 47.1%; Score 41; DB 2; Length 404;  
 Best Local Similarity 50.0%; Pred. No. 90;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 SYOHDLRAYGFV 13  
 :|:| | | |  
 DB 268 AVEHKERFYGV 279

RESULT 13  
 09X7J2 PRELIMINARY: PRT: 425 AA.

ID 09X7J2  
 AC 09X7J2:  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE TRANSPOSASE.  
 OS Pseudomonas stutzeri (Pseudomonas perfectomarina).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ML;  
 RX MEDLINE=99240403; PubMed=10223973;  
 RA Bolognese F., Di Lecce C., Galli E., Barbieri P.;

RT "Activation and inactivation of Pseudomonas stutzeri methylbenzene  
 RT catabolism pathways mediated by a transposable element.";  
 RL Appl. Environ. Microbiol. 65:1876-1882(1999).  
 DR EMBL: AJ012352; CAB42636.1; -;  
 DR InterPro: IPR002560; Transposase\_12.  
 DR Pfam: PF01610; Transposase\_12; 1.  
 SQ SEQUENCE 425 AA: 49494 MW; 4EE338A7465586BF CRC64;

Query Match 47.1%; Score 41; DB 2; Length 425;  
 Best Local Similarity 50.0%; Pred. No. 95;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 SYOHDLRAYGFV 13  
 :|:| | | |  
 DB 289 AVEHKERFYGV 300

RESULT 14  
 P78875 PRELIMINARY: PRT: 817 AA.

ID P78875  
 AC P78875:  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE TREHALOSE-6-PHOSPHATASE (EC 3.1.3.12) (TREHALOSE 6-PHOSPHATE  
 DE PHOSPHATASE) (TPP).  
 GN TPP1 OR SPAC19612.15c.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Canasado J., Vicente J., Soto T., Franco A., Castillo R., Gacto M.;  
 RT "Isolation and characterization of trehalose-6p phosphatase disrupted  
 RT mutants from Schizosaccharomyces pombe.";  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Oliver K., Harris D., Wood V., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 505-817 FROM N.A.  
 RC STRAIN=PR745;  
 RX MEDLINE=98162722; PubMed=9501991;  
 RA Yoshiohara S., Kato K., Nakai K., Okayama H., Nojima H.;  
 RT "Identification of open reading frames in Schizosaccharomyces pombe  
 RT cDNAs.";  
 RL DNA Res. 4:363-369(1997).  
 CC -1- CATALYTIC ACTIVITY: TREHALOSE 6-PHOSPHATE + H(2)O = TREHALOSE +  
 CC ORTHOPHOSPHATE.  
 CC -1- SIMILARITY: CONTAINS A DOMAIN FOUND IN ALL THE PROTEINS OF THE  
 CC TPS COMPLEX.  
 CC EMBL: AJ242743; CAB45142.1; -;  
 DR EMBL: 297209; CAB10126.1; -;  
 DR EMBL: D89225; BAB13886.1; -;  
 DR InterPro: IPR001830; Glyco\_transf\_20.  
 DR InterPro: IPR003337; Trehalose\_Pfase.  
 DR Pfam: PF00982; Glyco\_transf\_20; 1.  
 DR Pfam: PF02358; Trehalose\_Pfase; 1.  
 KM Hydrolase.  
 FT DOMAIN 1 547 TPS COMPLEX DOMAIN.  
 SQ SEQUENCE 817 AA: 93877 MW; 23CFD253505CBA CRC64;

Query Match 47.1%; Score 41; DB 3; Length 817;  
 Best Local Similarity 57.1%; Pred. No. 19e+02;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;



OS Leishmania major.  
 OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Borzym K., Klages S., Reinhardt R., Beck A., Ivens A.C., Quail M.,  
 Rajandream M.A., Barrell B.G.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 MEDLINE=98146435; PubMed=9477341;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome.";  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL; AL13618; CAB71284.2; -.  
 KW Hypothetical protein.  
 SO SEQUENCE 306 AA; 34129 MW; D03DA12477200AE7 CRC64;

Query Match 47.7%; Score 41.5; DB 5; Length 306;  
 Best Local Similarity 66.7%; Pred. NO. 55;  
 Matches 8; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 4 OHDLRAYGFNRL 15  
 ||: |||||  
 Db 213 QHN---YGFNRL 221

RESULT 9  
 Q9H356 PRELIMINARY; PRT; 61 AA.  
 AC Q9H356;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE PRO1804.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER.  
 RA Zhang C., Yu Y., Zhang S., Zhou G., Wei H., Bi J., Dong C., Zai Y.,  
 Xu W., Gao F., Liu M., He F.;  
 RT "Functional prediction of the coding sequences of 11 new genes deduced  
 by analysis of cDNA clones from human fetal liver.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF132201; AAC35547.1; -.  
 SO SEQUENCE 61 AA; 6678 MW; F238B0E082E080C9 CRC64;

Query Match 47.1%; Score 41; DB 4; Length 61;  
 Best Local Similarity 58.3%; Pred. NO. 11;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 SYOHLRAYGFN 13  
 |||: |||||  
 Db 6 SYONSLGLGFW 17

RESULT 10  
 Q9KT60 PRELIMINARY; PRT; 188 AA.  
 AC Q9KT60;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE PUTATIVE RNA POLYMERASE SIGMA FACTOR.  
 GN VCI045.

OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES  
 ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND  
 THEN IS RELEASED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.  
 DR EMBL; AE004186; AAF94204.1; -.  
 DR TIGR; VCI045; -.  
 DR InterPro; IPR000838; Sigma70\_ECF.  
 DR Pfam; PF00776; Sigma70\_ECF.1.  
 DR PROSITE; PS01063; SIGMA70\_ECF; UNKNOWN.1.  
 KW Complete proteome; DNA-binding; DNA-directed RNA polymerase;  
 KW Sigma factor; Transcription regulation.  
 SO SEQUENCE 188 AA; 22547 MW; 313FA8201085A2 CRC64;

Query Match 47.1%; Score 41; DB 16; Length 188;  
 Best Local Similarity 46.2%; Pred. NO. 39;  
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSYOHLRAYGFN 13  
 |:| || | | |  
 Db 32 RAYRDLRYAYW 44

RESULT 11  
 Q9Z5V9 PRELIMINARY; PRT; 338 AA.  
 AC Q9Z5V9;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE PUTATIVE TRANSPOSASE A.  
 GN TNP.A.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=142;  
 RA MEDLINE=99240444; PubMed=10224014;  
 RX Tsol T.V., Plotnikova E.G., Cole J.R., Guerin W.F., Bagdasarian M.,  
 Tiedje J.M.;  
 RT "Cloning, expression, and nucleotide sequence of the Pseudomonas  
 aeruginosa 142 ohb genes coding for oxygenolytic ortho dehalogenation  
 of haloaromatics.";  
 RL Appl. Environ. Microbiol. 65:2151-2162(1999).  
 DR EMBL; AF121970; AAD20008.1; -.  
 DR InterPro; IPR002560; Transposase\_12.  
 DR Pfam; PF01610; Transposase\_12; 1.  
 SO SEQUENCE 338 AA; 39711 MW; E0C3DF2EAB3B329 CRC64;

Query Match 47.1%; Score 41; DB 2; Length 338;  
 Best Local Similarity 50.0%; Pred. NO. 74;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

	Matches	9; Conservative	3;	Mismatches	3;	Indels	6;	Gaps	1
QY	1	RSYQHDLRA-----YGFRL	15	:	:				
Db	189	REYQUNLSLARVEIYGYWRV	209						

RESULT	6			
ID	09VJH0	PRELIMINARY:	PRT:	652 AA.
AC	09VJH0:			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DE	CG13283 PROTEIN.			
GN	CG13283.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY.			
RA	MEDLINE=20196006; PubMed=10731132.			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sultion G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chape M., Pfeiffer B.D.,			
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Arril J.F., Agayanni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Bailly R.M., Basu A., Baxendale J., Bayraktarov I., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Duran-Rocha S., Dunkov B.C., Dunn P.,			
RA	Dubin R.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,			
RA	Fosler C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,			
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,			
RA	Hoshti D., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,			
RA	Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merklow G., Mishina N.V., Mobarry C., Morris J., Mosher A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy C., Morris J., Mosher A.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Schmeier F., Smith H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smet T.,			
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Massarone D.A., Wehnstock G.M., Weissbach J.,			
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,			
RT	"The genome sequence of Drosophila melanogaster."			
RL	Science 287:2185-2195(2000).			
DR	EMBL: AEO03652; AAF53579.1; -			
DR	PLAYbase: FBgn0032613; CG13283.			
QO	SEQUENCE 652 AA: 77836 MW: 8045549AB2D955A2 CRC64:			

Query Match	48.3%	Score 42	DB 5	Length 652
Best Local Similarity	53.8%	Pred. No. 1e+02		
Matches	7	Conservative	2	Mismatches 4
				Indels 0
				Gaps 0
xy	1	RSYQHDLRAYGF	13	

```

Db      557  RSFQYSLTDYHEW 569
          ||:|:|  ||

```

RESULT	7		
ID	004242	PRELIMINARY;	PRT: 887 AA.
AC	004242;		
DT	01-JUL-1997 (TREMBLrel. 04, Created)		
DT	01-JUL-1997 (TREMBLrel. 04, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	POTASSIUM CHANNEL.		
GN	ZKRL.		
OS	Zea mays (Maize).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;		
OC	Panicoidae; Andropogoneae; Zea.		
OX	NCBI_TaxId=4577;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=L55131B, AND CV. APACHE; TISSUE=COLEOPTILE;		
RX	MEDLINE=97272307; Pubmed=9114073;		
RA	Hoth S., Dreyer I., Dietrich P., Becker D., Mueller-Roeber B.,		
RA	Hedrich R.;		
RT	"Molecular basis of plant-specific acid activation of K+ uptake		
RT	channels.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 94:4806-4810(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=L55131B, AND CV. APACHE; TISSUE=COLEOPTILE;		
RX	MEDLINE=20604478; Pubmed=10518597;		
RA	Philippart K., Fuchs I., Luthen H., Hoth S., Bauer C.S., Haga K.,		
RA	Thiel G., Jung K., Sandberg G., Botter M., Becker D., Hedrich R.;		
RT	"Auxin-induced K+ channel expression represents an essential step in		
RT	coleoptile growth and gravitropism.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 96:12186-12191(1999).		
DR	EMBL; Y07633; CAA68912.1; -.		
DR	HSSP; 000420; IAWC.		
DR	InterPro: IPR002110; ANK		
DR	InterPro: IPR000636; Cation_chan_non_11g.		
DR	InterPro: IPR001622; Channel_pore_K.		
DR	InterPro: IPR000595; CNMP_binding.		
DR	Pfam: PF00023; ank; 5.		
DR	Pfam: PF00027; CNMP_binding; 1.		
DR	Pfam: PF00520; ion_trans; 1.		
DR	PRINTS; PR01415; ANKYRIN.		
DR	SMART; SM00248; ANK; 2.		
DR	SMART; SM00100; CNMP; 1.		
DR	PROSITE; PS50088; ANK_REPEAT; 2.		
DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.		
DR	PROSITE; PS50042; CNMP_BINDING_3; 1.		
DR	ANK repeat; Repeat.		
SD	SEQUENCE 887 AA; 99265 MW; D85605200A6B419D CRC64;		

Query Match	48.3%	Score 42;	DB 10;	Length 887;
Best Local Similarity	56.2%	Pred. No. 1.5e+02;		
Matches	9;	Conservative	1;	Mismatches 0;
				Indels 6;
				Gaps 1.
Oy	6	DLRATG-----RRL	15	
		:		
		:		
		:		
Db	163	DLRSTGFNNMLRL	178	
RESULT	8			
O9NEGA				
ID	O9NEGA	PRELIMINARY;	PRT;	306 AA.
AC	O9NEGA			
DT	01-OCT-2000	(TREMBLrel. 15, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	HYPOTHETICAL 34.1	KDA PROTEIN.		
LN	L2954.12.			

DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE NADH DEHYDROGENASE 1, A SUBUNIT.  
 GN NMB0241.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J.,  
 Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,  
 Cotton M.D., Ulfenbach T.R., Khouri H., Qin H., Vamathevan J.,  
 Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,  
 Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58."  
 RL Science 287:1809-1815(2000).  
 DR EMBL: AE002381; AAF40695.1; -  
 DR TIGR: NMB0241; -  
 DR InterPro: IPR000440; Oxidored\_q4.  
 DR Pfam: PF00507; Oxidored\_q4; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 118 AA; 13598 MW; 21F58B3BAC579F84 CRC64;

Query Match 49.4%; Score 43; DB 16; Length 118;  
 Best Local Similarity 70.0%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DLDRYGFWR 15  
 || ||||| :  
 Db 86 DLGAYGFSM 95

RESULT 3  
 Q9JX78 PRELIMINARY; PRT; 118 AA.  
 AC Q9JX78;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE NADH DEHYDROGENASE I CHAIN A (EC 1.6.5.3).  
 GN NUO OR NMA0019.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=56599;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
 Jorgels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 Whitehead S., Spratt B.G., Barrall B.G.,  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 22491."  
 RL Nature 404:502-506(2000).  
 DR EMBL: AL162752; CAB83339.1; -  
 DR InterPro: IPR000440; Oxidored\_q4.  
 DR Pfam: PF00507; Oxidored\_q4; 1.  
 KW Oxidoreductase; Complete proteome.  
 SQ SEQUENCE 118 AA; 13584 MW; 38C401130EDDAE17 CRC64;

Query Match 49.4%; Score 43; DB 16; Length 118;  
 Best Local Similarity 70.0%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DLDRYGFWR 15  
 || ||||| :  
 Db 86 DLGAYGFSM 95

RESULT 4  
 Q25352 PRELIMINARY; PRT; 188 AA.  
 ID Q25352  
 AC Q25352; 002505;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE I2759.14.  
 GN I2759.14.  
 OS Leishmania major.  
 OC Eukaryota; Eucaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIDLIN;  
 RX MEDLINE=99178987; PubMed=10077609;  
 RA Myler P.J., Audleman L., deVos T., Hixson G., Kiser P., Lemley C.,  
 Magness C., Rickel E., Sisk E., Sunkin S., Swartzell S., Westlake T.,  
 Bastien P., Fu G., Ivens A., Stuart K.,  
 RT "Leishmania major Friedlin chromosome 1 has an unusual distribution of  
 protein-coding genes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2902-2906(1999).  
 DR EMBL: AE001274; AAC24644.1; -  
 SQ SEQUENCE 188 AA; 20964 MW; EE4A9F2B76E093B CRC64;

Query Match 49.4%; Score 43; DB 5; Length 188;  
 Best Local Similarity 77.8%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 HDLDRYGFWR 13  
 ||| | ||| :  
 Db 46 HDLMAIGFW 54

RESULT 5  
 Q22405 PRELIMINARY; PRT; 444 AA.  
 ID Q22405  
 AC Q22405;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE T11F9.1 PROTEIN.  
 GN T11F9.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloiderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lennard N.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z74042; CA98536.1; -  
 DR InterPro: IPR000276; GPCR\_Rhodopsn.  
 DR PROSITE: PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
 SQ SEQUENCE 444 AA; 51686 MW; 3C07F095C5484892 CRC64;

Query Match 48.3%; Score 42; DB 5; Length 444;  
 Best Local Similarity 42.9%; Pred. No. 68;



CC or send an email to license@slb.ch).  
 CC -----  
 DR EMBL; U91738; AAB58393.1; -  
 DR InterPro; IPR001370; BIR.  
 DR Pfam; PF00653; BIR; 1.  
 DR SMART; SM00238; BIR; 1.  
 DR PROSITE; PS01282; BIR\_REPEAT\_1; 1.  
 DR PROSITE; PS50143; BIR\_REPEAT\_2; 1.  
 FT REPEAT 43 106 BIR.  
 SQ SEQUENCE 238 AA; 28530 MW; 7024C187D2C55E40 CRC64;

Query Match 43.7%; Score 38; DB 1; Length 238;  
 Best Local Similarity 50.0%; Pred. No. 19;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 QHDLRAYGFW 13  
 : 1 1 : 111  
 Db 90 KHPLOYHGFW 99

Search completed: August 15, 2002, 11:47:41  
 Job time: 517 sec

```

OY      4 QHDLRAYGFW 13
      : 1 1 : 1111
      76 KHPLOYHGFW 85

RESULT 13
ID      IAPL_ASFC3      STANDARD:      PRT:      224 AA.
AC      011451:
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      IAP-1-like protein p27.
GN      4CL.
OS      African swine fever virus (isolate crocodile/96/3 / C-3) (ASFV).
OC      Viruses: dsDNA viruses, no RNA stage; Asfarviridae.
OC      African swine fever-like viruses.
OX      NCBI_TaxId=82811;

RN      SEQUENCE FROM N.A.
RP      MEDLINE=97288303; PubMed=9143281;
RA      Neilan J.G., Lu Z., Kutish G.F., Zsak L., Burrage T.G.,
RA      Borca M.V., Carrillo C., Rock D.L.;
RT      "A BIR motif containing gene of African swine fever virus, 4CL, is
RT      nonessential for growth in vitro and viral virulence."
RL      Virology 230:252-264(1997).
CC      -1- FUNCTION: NOT ESSENTIAL FOR GROWTH OR VIRULENCE. DOES NOT HAVE
CC      ANTIAPOPTOTIC FUNCTION.
CC      -1- SIMILARITY: CONTAINS 1 BIR REPEAT.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL: U91732; AAB58387.1; -
DR      InterPro: IPR001370; BIR.
DR      Pfam: PF00653; BIR.1.
DR      SMART: SM00238; BIR.1.
DR      PROSITE: PS01282; BIR_REPEAT_1; 1.
DR      PROSITE: PS50143; BIR_REPEAT_2; 1.
FT      REPEAT      29      92      BIR.
SQ      SEQUENCE      224 AA; 26559 MW; 80B376948514946F CRC64;

Query Match      43.7%; Score 38; DB 1; Length 224;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY      4 QHDLRAYGFW 13
      : 1 1 : 1111
      76 KHPLOYHGFW 85

RESULT 14
ID      IAPL_ASFC3      STANDARD:      PRT:      224 AA.
AC      012407:
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      IAP-1-like protein p27.
GN      4CL.
OS      African swine fever virus (isolate Chiredzi/83/1 / Ch1) (ASFV),
OS      African swine fever virus (isolate crocodile/96/1 / Crl) (ASFV), and
OS      African swine fever virus (isolate Pretoriuskop/96/5 / P-5) (ASFV).
OC      Viruses: dsDNA viruses, no RNA stage; Asfarviridae.
OC      African swine fever-like viruses.
OX      NCBI_TaxId=82812, 82810, 82817;

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97288303; PubMed=9143281;
RA      Neilan J.G., Lu Z., Kutish G.F., Zsak L., Burrage T.G., Borca M.V.,
RA      Carrillo C., Rock D.L.;
RT      "A BIR motif containing gene of African swine fever virus, 4CL, is
RT      nonessential for growth in vitro and viral virulence."
RL      Virology 230:252-264(1997).
CC      -1- FUNCTION: NOT ESSENTIAL FOR GROWTH OR VIRULENCE. DOES NOT HAVE
CC      ANTIAPOPTOTIC FUNCTION.
CC      -1- SIMILARITY: CONTAINS 1 BIR REPEAT.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL: U91730; AAB58385.1; -
DR      EMBL: U91731; AAB58386.1; -
DR      EMBL: U91740; AAB58395.1; -
DR      InterPro: IPR001370; BIR.
DR      Pfam: PF00653; BIR.1.
DR      SMART: SM00238; BIR.1.
DR      PROSITE: PS01282; BIR_REPEAT_1; 1.
DR      PROSITE: PS50143; BIR_REPEAT_2; 1.
FT      REPEAT      29      92      BIR.
SQ      SEQUENCE      224 AA; 26558 MW; 8E579C948F132F05 CRC64;

```

```

OY      4 QHDLRAYGFW 13
      : 1 1 : 1111
      76 KHPLOYHGFW 85

RESULT 15
ID      IAPL_ASFC2      STANDARD:      PRT:      238 AA.
AC      011453:
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      IAP-1-like protein p27.
GN      4CL.
OS      African swine fever virus (isolate Malawi L11 20/1) (ASFV).
OC      Viruses: dsDNA viruses, no RNA stage; Asfarviridae.
OC      African swine fever-like viruses.
OX      NCBI_TaxId=10500;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-MALAWI;
RX      MEDLINE=97288303; PubMed=9143281;
RA      Neilan J.G., Lu Z., Kutish G.F., Zsak L., Burrage T.G.,
RA      Borca M.V., Carrillo C., Rock D.L.;
RT      "A BIR motif containing gene of African swine fever virus, 4CL, is
RT      nonessential for growth in vitro and viral virulence."
RL      Virology 230:252-264(1997).
CC      -1- FUNCTION: NOT ESSENTIAL FOR GROWTH OR VIRULENCE. DOES NOT HAVE
CC      ANTIAPOPTOTIC FUNCTION.
CC      -1- SIMILARITY: CONTAINS 1 BIR REPEAT.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



Query Match 44.8%; Score 39; DB 1; Length 700;  
 Best Local Similarity 50.0%; Pred. No. 40;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 RSYQHDRLRAYGFW 14  
 : | | : | |  
 DB 53 RSMQAEIRLTFEWR 66

RESULT 10  
 VGLM\_TSWV1 STANDARD: PRT; 1135 AA.  
 AC P36291;

DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE M polypeptide precursor [Contains: Glycoprotein G1; Glycoprotein G2].  
 GN M.  
 OS Tomato spotted wilt virus (strain Brazilian Br-01) (TSWV).  
 OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.  
 OX NCBI\_TaxID=36413;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93057351; PubMed=1431808;  
 RA Kormelink R., de Haan P., Meurs C., Peters D., Goldbach R.;  
 RT "The nucleotide sequence of the M RNA segment of tomato spotted wilt virus, a bunyavirus with two ambisense RNA segments.";  
 RL J. Gen. Virol. 73:2795-2804(1992).  
 RN [2]  
 RP ERRATUM.  
 RX MEDLINE=93224906; PubMed=8468562;  
 RA Kormelink R., de Haan P., Meurs C., Peters D., Goldbach R.;  
 RL J. Gen. Virol. 74:790-790(1993).

CC -1 SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (Probable).  
 CC -1 PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD NATURE PROTEINS INCLUDING GLYCOPROTEIN G1 AND GLYCOPROTEIN G2.  
 CC -1 SIMILARITY: BELONGS TO THE HANTAVIRUSES M POLYPEPTIDE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; S48091; AAB24089.1; -;  
 DR PIR; J01928; J01928.  
 KW Polypeptide; Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 36 1135 M POLYPEPTIDE.  
 FT CHAIN ? 1135 GLYCOPROTEIN G1.  
 FT SITE 41 43 CELL ATTACHMENT SITE (POTENTIAL).  
 FT TRANSMEM 315 335 POTENTIAL.  
 FT TRANSMEM 346 366 POTENTIAL.  
 FT TRANSMEM 437 457 POTENTIAL.  
 FT TRANSMEM 749 769 POTENTIAL.  
 FT TRANSMEM 1068 1088 POTENTIAL.  
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 605 605 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1095 1095 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 1135 AA; 127317 MW; AC1A3FFEB4044FB CRC64;

Query Match 44.8%; Score 39; DB 1; Length 1135;

Best Local Similarity 46.2%; Pred. No. 66;  
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 RSYQHDRLRAYGFW 13  
 : | | : | |  
 DB 1033 QOYNSSDSAWGFW 1045

RESULT 11  
 OMPX\_ECOLI STANDARD: PRT; 171 AA.  
 AC P36546;

DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein X precursor.  
 GN OMPX OR B0814 OR Z1036 OR ECS0892.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=95138044; PubMed=7836315;  
 RA Mecsas J., Welch R., Erickson J.W., Gross C.A.;  
 RT "Identification and characterization of an outer membrane protein, OmpX, in Escherichia coli that is homologous to a family of outer membrane proteins including Ail of Yersinia enterocolitica.";  
 RL J. Bacteriol. 177:799-804(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Alpha H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-153(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,



DR	EMBL:	AEO00360;	AACT5808.1;	ALT_INIT.
DR	EcoGene:	EGJ3121;	Y9CN	
DR	InterPro:	IPR000205;	NAD_binding.	
KW	Hypothetical protein:	Oxidoreductase:	FAD; Flavoprotein;	
KW	Complete proteome.			
FT	NP_BIND	7	21	FAD (ADP PART) (POTENTIAL).
SQ	SEQUENCE	423 AA;	46393 MW;	346ED1A132F31DEC CRC64;
Query Match	Best Local Similarity	44.8%;	Score 39;	DB 1; Length 423;
Matches	6; Conservative	2;	Mismatches	5; Indels 0; Gaps 0;
Oy	3 YOHDLRAFGFWRL 15			
Db	400 FWHHLRRHGLMHL 412			
RESULT	9			
BIB_DROME	BIB_DROME	STANDARD;	PRT;	700 AA.
ID	P33645;			
AC	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Neurogenic Protein big brain.			
GN	BIB.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
CC	Empidoidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90245091; PubMed=1692392;			
RA	Rao Y., Jan L.Y., Jan Y.N.;			
RT	"Similarity of the product of the Drosophila neurogenic gene big			
RT	brain to transmembrane channel proteins.";			
RL	Nature 345:163-167(1990).			
CC	-I- FUNCTION: MAY MEDIATE INTERCELLULAR COMMUNICATION; MAY FUNCTIONS			
CC	BY ALLOWING THE TRANSPORT OF CERTAIN MOLECULE(S) AND THEREBY			
CC	SENDING A SIGNAL FOR AN EXODERMAL CELL TO BECOME AN EPIDERMAL			
CC	INSTEAD OF A NEUROBLAST. MUTATION IN BIB GENE UNDERLIES "BIG			
CC	BRAIN DEVELOPMENT.			
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-I- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/sib.ch).			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL:	X53775;	CAB37863.1;	-
DR	PIR:	S09699;	S09699.	
DR	FLYBase:	FBgn0000180;	bib.	
DR	InterPro:	IPR000425;	MIP.	
DR	Pfam:	PF00230;	MIP; 1.	
DR	PRINTS:	PR00783;	MIMTRINSICP.	
DR	PROSITE:	PS00221;	MIP; 1.	
KW	Developmental protein;	Neurogenesis;	Transport;	Transmembrane.
FT	TRANSMEM	72	93	POTENTIAL.
FT	TRANSMEM	98	118	POTENTIAL.
FT	TRANSMEM	144	167	POTENTIAL.
FT	TRANSMEM	190	208	POTENTIAL.
FT	TRANSMEM	214	234	POTENTIAL.
FT	TRANSMEM	257	273	POTENTIAL.
FT	DOMAIN	439	451	POLY-GLN.
FT	DOMAIN	487	534	GLN-RICH.
FT	DOMAIN	658	690	POLY-GLN.
SO	SEQUENCE	700 AA;	76951 MW;	3390A6B3F96D5B6B CRC64;

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rongning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umeyam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Cressy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
CC -I- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity)
CC -I- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ZETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AC005312; AAC78520.1; -
DR Mendel: 38453; Arath:1313;38453.
DR InterPro: IPR004046; GST_C.
DR InterPro: IPR004045; GST_N.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR Transferase.
KM ACT SITE 20
FT SEQUENCE 223 AA; 25352 MW; AAE7E85746610512 CRC64;
SQ
Query Match 46.0%; Score 40; DB 1; Length 223;
Best Local Similarity 50.0%; Pred. No. 8;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 3 YQHDRLAYGFWR 14
DB 8 YQAKLKLYSYMR 19
RESULT 5
VOBL_CAEEL STANDARD: PRT; 451 AA.
AC 009255;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 51.0 kDa protein C30612.1 in chromosome II.
GN C30612.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RA Lareille P.;
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: U21319; AAC46670.1; -
DR WormPep: C30612.1; CE01832.
KM Hypothetical protein.
SQ SEQUENCE 451 AA; 51017 MW; 6AAAFD326751C1 CRC64;
Query Match 46.0%; Score 40; DB 1; Length 451;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 SYQHDRLAY 10
DB 61 SYQHDLSAF 69
RESULT 6
YIHX_ECOLI STANDARD: PRT; 199 AA.
ID YIHX_ECOLI
AC P32145;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yihx.
GN YIHX OR B385.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=93347969; PubMed=8346018;
RA Plunkett G., III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes."
RL Nucleic Acids Res. 21:3391-3398(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L19201; AAB03018.1; ALT_INIT.
DR EMBL: AE000464; AAD13447.1; ALT_INIT.
DR PIR: S40829; S40829.
DR Ecogene: EG11850; yihx.
DR InterPro: IPR001454; Hydrolase.
DR Pfam: PF00702; Hydrolase; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 199 AA; 22732 MW; 9451EA891DFC0D2 CRC64;
Query Match 44.8%; Score 39; DB 1; Length 199;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 RSYQHDRLAYGF 12
DB 145 RIYQHVLAQEGF 156
RESULT 7
VG44_HSV1 STANDARD: PRT; 335 AA.
ID VG44_HSV1
AC 000124;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Hypothetical gene 44 protein.
GN 44.

```

FT TRANSMEM 423 443 POTENTIAL.  
SQ SEQUENCE 459 AA; 49941 MW; B2EPD15C8EA94E CRC64;

Query Match 49.4%; Score 43; DB 1; Length 459;  
Best Local Similarity 50.0%; Pred. No. 5.2;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 SYQHDLRAYGFW 13  
: | : | | |  
Db 416 AYRFDMGITGFW 427

RESULT 2  
SYLM\_NEUCR STANDARD; PRT; 994 AA.  
ID SYLM\_NEUCR

AC P5181;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Leucyl-tRNA synthetase, mitochondrial precursor (EC 6.1.1.4)  
DE (Leucine--tRNA ligase) (LeuRS).  
GN LEU-5 OR 99H12.170.

OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90097874; PubMed=2574823;

RA Chow C.M., Metzgerberg R.L., Rajbhandary U.L.;

RT "Nuclear gene for mitochondrial leucyl-tRNA synthetase of Neurospora crassa: isolation, sequence, chromosomal mapping, and evidence that

the leu-5 locus specifies structural information.";  
RN Mol. Cell. Biol. 9:4631-4644(1989).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=OR74A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.-W., Mannhaupt G.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +

CC diphosphate + L-leucyl-tRNA(Leu).

CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.

CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: M30472; AAA33599.1; -

DR EMBL: AL451018; CAC18253.1; -

DR PIR: A33474; SYNCML.

DR InterPro: IPR002300; tRNA-synt\_1a.

DR InterPro: IPR001412; tRNA-synt\_1.

DR InterPro: IPR002302; tRNA-synt\_1eu.

DR Pfam: PF00133; tRNA-synt\_1.1.

DR PRINTS: PR00985; TRNASYNTH1EU.

DR PROSITE: PS00178; AA-tRNA-LIGASE\_1; 1.

KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW Mitochondrion; Transit peptide.

FT TRANSLAT 1 52 MITOCHONDRION.

FT CHAIN 53 994 LEUCYL-tRNA SYNTHETASE.

FT SITE 93 103 "HIGH" REGION.

FT STATE 697 701 "KMSKS" REGION.

FT BINDING 700 700 ATP (BY SIMILARITY).

FT SEQUENCE 994 AA; 109853 MW; 2ADB26302A8465F9 CRC64;

Query Match 48.3%; Score 42; DB 1; Length 994;  
Best Local Similarity 60.0%; Pred. No. 17;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 HDLRAYGFW 14  
| | | | : | |  
Db 403 HDLRDHAFWK 412

RESULT 3  
RL5\_NEUCR STANDARD; PRT; 301 AA.  
ID RL5\_NEUCR

AC O59953;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 60S ribosomal protein L5 (CPR4).

GN CRP-4.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=74-OR23-1A.

RA de la Serna I.L., Cujec T.P., Shi Y., Tyler B.M.;

RT "Non-coordinate regulation of 5S rRNA genes and the gene encoding the

5S rRNA-binding ribosomal protein homolog in Neurospora crassa.";

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: THIS PROTEIN BINDS 5S RNA (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AF054907; AAC09000.1; -

DR InterPro: IPR001149; Ribosomal\_L18p.

DR Pfam: PF00861; Ribosomal\_L18p.1.

DR PRINTS: PR00058; RIBOSOMAL15.

DR PRODOM: PD001394; Ribosomal\_L18p.1.

KW Ribosomal protein; RNA-binding.

SQ SEQUENCE 301 AA; 34411 MW; 289B0F0D1672A5F9 CRC64;

Query Match 47.1%; Score 41; DB 1; Length 301;  
Best Local Similarity 60.0%; Pred. No. 7.4;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SYQHDLRAYG 11  
: | : | | : | | |  
Db 78 AYSHELRAYG 87

RESULT 4  
GT2L\_ARATH STANDARD; PRT; 223 AA.  
ID GT2L\_ARATH

AC O92VQ4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable glutathione S-transferase zeta-class 1 (EC 2.5.1.18).

GN AT2G03380 OR T16F16.17.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1]

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:47:41 ; Search time 18.67 Seconds

(without alignments)  
31.108 Million cell updates/sec

Title: US-09-613-092a-6

Perfect score: 87

Sequence: 1 RSYQHDLRAYGFWRLL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues 105224

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	49.4	459	1	NORM_NEIMA
2	42	48.3	994	1	SYM_NEUCR
3	41	47.1	301	1	RIS_NEUCR
4	40	46.0	223	1	GT2L_ARATH
5	40	46.0	451	1	YOB1_CAEEL
6	39	44.8	199	1	YIH1_ECOLI
7	39	44.8	335	1	VG44_HSV1
8	39	44.8	423	1	YGCN_ECOLI
9	39	44.8	700	1	BIB_DROME
10	39	44.8	1135	1	VG1M_TSW1
11	38	43.7	171	1	OMPX_ECOLI
12	38	43.7	224	1	IAPL_ASFB7
13	38	43.7	224	1	IAPL_ASFC3
14	38	43.7	224	1	IAPL_ASFC4
15	38	43.7	238	1	IAPL_ASFC4
16	38	43.7	259	1	IAPL_ASFC4
17	38	43.7	459	1	NORM_NEIMA
18	38	43.7	978	1	YX77_CAEEL
19	38	43.7	1192	1	LMG2_MOUSE
20	38	43.7	1193	1	LMG2_MOUSE
21	37	42.5	142	1	ELBS_ADE04
22	37	42.5	319	1	EXOM_RHIME
23	37	42.5	331	1	YJ43_SYNZ3
24	37	42.5	453	1	OP2_MAIZE
25	37	42.5	514	1	AMT2_LYCES
26	37	42.5	667	1	GR78_APICA
27	37	42.5	715	1	PERE_HUMAN
28	37	42.5	716	1	PERE_MOUSE
29	37	42.5	745	1	PERM_MOUSE
30	37	42.5	877	1	AGLU_HOVRT
31	37	42.5	941	1	CHRD_XENLA
32	37	42.5	999	1	MPD_HELPY
33	37	42.5	1001	1	MPD_HELPY

34	37	42.5	1613	1	VIR2_CAEEL	P05690 caenorhabd
35	37	42.5	1616	1	VIR1_CAEEL	P55155 caenorhabd
36	37	42.5	2329	1	YJ16_CAEEL	P34369 caenorhabd
37	36	41.4	68	1	Y002_BHP1	P51701 bacterioph
38	36	41.4	97	1	YA9_HAETN	O57347 haemophilus
39	36	41.4	301	1	Y186_MYCPN	P75265 mycoplasma
40	36	41.4	331	1	HPIC_TREPA	O83152 treponema
41	36	41.4	347	1	YPPG_ECOLI	P76559 escherichia
42	36	41.4	374	1	YLD1_CAEEL	O03566 caenorhabd
43	36	41.4	403	1	YBL2_SFVL	P29170 simian foam
44	36	41.4	559	1	Y876_MYCTU	O10564 mycobacteri
45	36	41.4	571	1	YR10_CAEEL	P90794 caenorhabd

## ALIGNMENTS

RESULT	ID	NAME	STANDARD	PRT	AA
AC	09J27	NORM_NEIMA			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Probable multidrug resistance protein norm (Na(+)/drug antiporter)				
DE	(Multidrug-efflux transporter).				
GN	NORM OR NMA1022.				
OS	Neisseria meningitidis (serogroup A).				
OC	Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=65699;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;				
RX	MEDLINE=20222556; PubMed=10761919;				
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,				
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,				
RA	Davies R.M., Davis P., Devlin K., Fellwell T., Hamlin N., Holtroyd S.,				
RA	Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,				
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,				
RA	Whitehead S., Spratt B.G., Barrell B.G.;				
RT	"Complete DNA sequence of a serogroup A strain of Neisseria				
RT	meningitidis Z2491."				
RL	Nature 404:502-506 (2000).				
CC	-1- FUNCTION: FUNCTIONS AS A NA(+)/DRUG ANTIporter (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane				
CC	(Potential).				
CC	-1- SIMILARITY: BELONGS TO THE MULTI ANTIMICROBIAL EXTRUSION (MATE)				
CC	FAMILY. NORM (TC 2.A.66.1.1) SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).				
CC	-----				
DR	EMBL: AL162754; CAB84291.1; "				
DR	InterPro: IPR002528; UPF0013.				
DR	Pfam: PF01554; UPF0013; 2.				
KW	Transport; Sodium transport; Transmembrane; Inner membrane;				
KW	Complete proteome.				
FT	TRANSMEM	20	40		POTENTIAL.
FT	TRANSMEM	53	73		POTENTIAL.
FT	TRANSMEM	100	120		POTENTIAL.
FT	TRANSMEM	132	152		POTENTIAL.
FT	TRANSMEM	168	188		POTENTIAL.
FT	TRANSMEM	202	222		POTENTIAL.
FT	TRANSMEM	252	272		POTENTIAL.
FT	TRANSMEM	285	305		POTENTIAL.
FT	TRANSMEM	325	345		POTENTIAL.
FT	TRANSMEM	358	378		POTENTIAL.
FT	TRANSMEM	395	415		POTENTIAL.